DE GRUYTER MOUTON Linguistics Vanguard 2022; 8(s1): 191–203

Jonas Nölle* and Michael Spranger

From the field into the lab: causal approaches to the evolution of spatial language

https://doi.org/10.1515/lingvan-2020-0007
Received February 2, 2020; accepted June 14, 2021

Abstract: Striking variation exists in preferences for specific spatial linguistic strategies among different speech communities. Increasing evidence now suggests that this might not simply be a result of neutral drift, but rather a form of linguistic adaptation to the local social, cultural, or physical environment. Recent studies indicate that different factors like topography, subsistence style, and bilingualism successfully predict the choice of spatial frames of reference (FoR) on linguistic and non-linguistic tasks. However, the exact causal relationships between these variables and the cultural evolutionary mechanisms behind the selection of one FoR strategy over another are still not fully understood. In this paper, we argue that to arrive at a more mechanistic and causal understanding of the cultural evolution of spatial language, observations from descriptive fieldwork should be combined with experimental and computational methods. In the framework we present, causal relationships between linguistic and non-linguistic variables (such as FoR choice and topography) can be isolated and systematically tested in order to shed light on how sociotopographic factors motivate the variation in spatial language we observe cross-linguistically. We discuss experimental results from behavioral studies and computer simulations that illustrate how this approach can deliver empirical findings that go beyond simple correlations.

Keywords: causality; frames of reference; language evolution; linguistic adaptation; spatial language

1 Introduction

Increasing empirical evidence suggests that language structure does not evolve in a void, but rather adapts to its wider social, cultural, and physical environment (see Lupyan and Dale [2016] for a review). Proposed links between linguistic and environmental variables include the relationships between morphological complexity and population structure (Bentz and Winter 2013; Lupyan and Dale 2010), between climate and sound systems (Everett et al. 2015), lexical categories (Brown and Lindsey 2004; Regier et al. 2016), and even subtle biases resulting from anatomy (Dediu et al. 2017), diet (Blasi et al. 2019), or communicative pressures (Coupé et al. 2019). One domain for which links between linguistic and non-linguistic variables remain hotly debated is spatial referencing, or more precisely, variation in the use of spatial frames of reference (FoR), conceptual coordinate systems used to express spatial relations between objects. For instance, certain speech communities prefer a viewpoint-centered “egocentric” FoR to relate a figure to a ground object (e.g., *the ball is to the left of the car*), while others prefer an absolute or environment-centered “geocentric” FoR (e.g., *the ball is north/uphill of the car*; see, e.g., Levinson 2003; Levinson and Wilkins 2006; Majid et al. 2004). It has been argued that these differences do not simply constitute cross-linguistic variation, but also affect speakers’ general judgments and cognitive strategies on non-linguistic spatial tasks (e.g., Haun et al. 2011; Levinson 2003; Majid et al. 2004; Pederson et al. 1998), although this remains highly controversial (e.g., Diessel 2014; Gallistel 2002; Li et al. 2011; Li and Gleitman 2002; Newcombe and Huttenlocher 2000; Pinker 2007). However, independently of the neo-Whorfian debate, it remains an unresolved issue where these systems originate from, and why they are distributed in the way are. A long philosophical tradition regarded egocentric spatial concepts as innate and

***Corresponding author: Jonas Nölle**, School of Psychology & Neuroscience, University of Glasgow, Glasgow, UK, E-mail: jonas.noelle@glasgow.ac.uk. https://orcid.org/0000-0003-3546-2689
Michael Spranger, Sony Computer Science Laboratories Inc, Tokyo, Japan. https://orcid.org/0000-0001-9443-7008

"natural", but this turned out to be a Eurocentric fallacy resulting from the lack of cross-cultural data (see, e.g., Levinson 2003: Ch. 1). Experimental evidence suggests that geocentric spatial reasoning might in fact be the default among primates and easier to acquire in development in both great apes (Haun et al. 2006) and human children (Shusterman and Li 2016). This suggests that geocentric FoR might be "older", while egocentric systems are a more recent innovation.

The overall variation and distribution of FoRs among the world's languages can be understood as the cumulative outcome of cultural evolutionary processes, which operate according to principles of function, fitness, replication, and selection (Dediu et al. 2013).[1] It is an open question whether this distribution is mainly due to neutral drift and historical contingency or can be explained in terms of adaptation to external variables such as the social or physical environment. In this paper, we argue that in order to answer this question, the relationship between spatial referencing and environment must be studied by looking through this cultural evolutionary lens. Spatial linguistic practices are cultural phenomena and as such deeply intertwined with other social or demographic factors. Cross-sectional observational data obtained from, for example, language description can serve as a starting point to describe variation, but correlations between linguistic and non-linguistic variables do not explain the underlying mechanisms. Instead, we suggest that causal links and dependencies can be made explicit and tested in isolation. Our aim is to demonstrate how this can be achieved with experiments that model the emergence and evolution of spatial referencing systems. Such experiments involve human participants or artificial agents and can address specific sociotopographic variables and their relationships to investigate causal effects and gain a more mechanistic understanding of FoRs and their origins and variation.

In Section 2, we will first review the challenges of traditional approaches to studying spatial linguistic phenomena, which largely result from the complex relationships between language and sociotopographic variables. We will show how causal graphs can help identify clear research questions by breaking up this complex web of interrelated variables into explicit causal hypotheses. Section 3 discusses how, in such a framework, evolutionary experiments can complement more traditional approaches by testing specific causal links that have been identified from observational data. We review a number of computational and behavioral experiments that have identified some of the potential mechanisms behind the spatial linguistic diversity observed in natural languages. These experiments demonstrate, for example, that it is possible to show the causal influence of single variables like "topography" on the selection of spatial referencing strategies. Furthermore, novel technologies, such as virtual reality, allow for the design of paradigms that elicit spatial language in very naturalistic or even large-scale settings, while simultaneously allowing for tight experimental control that enables the systematic testing of isolated effects. Such lab experiments can then be complemented with computational simulations that allow for the modeling of the evolution of spatial referencing systems over long time periods or in large populations of agents. The causal mechanisms found to shape spatial language in these experiments can then be verified against real-world data and thus uncover how spatial language culturally evolves in response to sociotopographic factors that lead to the striking variation we observe in the field.

Finally, we therefore conclude that such a "maximum robustness approach", including fieldwork, computational models, and lab experiments, is necessary to fully understand the relationship between spatial linguistic strategies and the environment in terms of their origins and underlying mechanisms.

2 Causality and the limitations of observational data

Speculation about environmental factors motivating FoR choice is as old as descriptive fieldwork documenting the diversity of spatial language: For instance, Wassmann and Dasen (1998) observed that geocentric Balinese spatial orientation systems shift along the coastline of the island's northeast peninsula in a way that appears to reflect the local topography. However, it has proven difficult to attribute such variation to a causal mechanism.

1 For an introduction to cumulative cultural evolution, see Richerson and Boyd (2008) and Mesoudi and Thornton (2018).

The first larger survey of twenty-one languages by Majid et al. (2004: 112) concluded that there is no discernible pattern which would suggest that environment or sociocultural factors such as dwelling or subsistence style correlate with a specific FoR.

It has since become clear that a major problem of earlier accounts was the equating of languages with specific FoRs; that is, using "language" rather than "speech community" or "field site" as the smallest unit of analysis. More recent comparative studies have shown that there is in fact a significant amount of variation within languages and speech communities if demographic factors are accounted for (Palmer et al. 2018). Factors which have been observed to correlate with specific FoR choices include age (Eggleston et al. 2011; Polian and Bohnemeyer 2011), first language (Bohnemeyer et al. 2015; Donelson 2018; Haun et al. 2011; Pederson et al. 1998), education or literacy (Danziger and Pederson 1998; Lin 2017; Mishra et al. 2003), gender (Ameka and Essegbey 2006; Bohnemeyer 2011; Lawton 2001; Le Guen 2011), occupation or subsistence style (Palmer et al. 2017; Shapero 2017), second language (Lin 2017; Moore 2018; Palmer 2015; Palmer et al. 2017), and various environmental factors such as settlement type (Adamou and Shen 2017; Mishra et al. 2003; Pederson et al. 1998) or local topography (Dasen and Mishra 2010; Palmer et al. 2017). Considering these factors within languages has revealed that even speakers of languages that were traditionally viewed as egocentric, such as Spanish (Adamou and Shen 2017; Bohnemeyer et al. this issue; Calderón et al. 2019), can prefer geocentric FoR on certain verbal or non-verbal tasks. This data comparing field sites rather than holistic "languages" suggests that variation in spatial language and cognition is deeply intertwined with culture and sociotopographic factors (see Lum et al. this issue). The main challenge is thus to disentangle how these different factors could shape variation over the course of cultural evolution as languages are used and transmitted to newer generations.

We suggest that modeling these cultural evolutionary processes and isolating explanatory variables is crucial to overcome some of the limitations of traditional descriptive approaches. For example, comparative fieldwork is costly and time-consuming. A survey of FoR use among thirteen Mesoamerican languages required a large-scale research project involving many collaborators (O'Meara and Pérez Báez 2011). Hence, many descriptions of spatial grammars stem from different research groups using different methods and even different theoretical frameworks, leading to differences in terms of, for example, how FoRs are defined and coded (Bohnemeyer et al. 2015: 175). This makes it hard to perform meta-analyses on data collected by different researchers, sometimes decades apart. Similarly, the complex nature of cultural phenomena involving many variables means that any new field site might reveal another previously unknown factor, which, in order to make a universal claim, should ideally be accounted for in all related studies, although this is often impossible post hoc. For instance, Palmer et al. (2017) developed the Sociotopographic Model from his earlier Topographic Correspondence Hypothesis (2015) after further studies revealed that social variables like subsistence style can modulate the effect of topography. The idiosyncratic nature and specific history of a cultural community can challenge previous theories: Contrary to predictions derived from previous research, Calderón et al. (2019) observe that Spanish monolinguals in Mexico make use of geocentric conceptualizations that seem to have been transmitted from the surrounding indigenous languages. Additionally, it is often difficult to compare physical environments, since linguists rarely record geospatial data and there is no simple categorical classification system of the earth's crust that can easily be correlated with specific FoR choices (Bohnemeyer 2016). Cross-sectional data sets further bear the risk of spurious correlations. The more variables we include in a statistical model, the more likely it is that some of them will spuriously appear to be associated with the outcome variable. Recent years have seen an increase in correlational studies, such as studies which link linguistic features to economic behavior (e.g., Chen 2013; Feldmann 2019; Kim et al. 2017). However, without proper theoretical motivation and without taking into account possible confounds, such as the problem that cultural datapoints are not independent from each other ("Galton's problem"; see Naroll 1961), such correlations can hardly be assumed to be robust (Roberts and Winters 2013).

For spatial language, we therefore suggest taking a more careful, robust, and causal approach, as laid out by Roberts (2018). Roberts contrasts a "maximum validity method" with a "maximum robustness method". The former means using the most relevant (or "valid") test to address a formulated hypothesis and accepting the result as the optimal evaluation. The maximum robustness approach, on the other hand, does not favor a specific analysis to support a theory. Instead, the same hypothesis is tested with various methodologies on different data sets controlling for different effects, and a result is only considered "robust" if evidence converges and the effect survives this assessment. The problem is that linguists tend to favor the maximum validity method, while cultural evolutionary processes usually involve "long chain[s] of causal connections that span many disciplines and large range of appropriate methodologies" (Roberts 2018: 2), which the maximum robustness approach is better suited to address.

Take for example Bohnemeyer et al.'s (2015) work on Mesoamerican languages, where different predictor variables (such as topography or L2-Spanish use) vary in whether they significantly explain variance in FoR usage depending on specifications of different generalized linear mixed effects models. While the effect of topography (which only came out significant in one of four models), for example, might not be strong enough to support claims about the environment affecting FoR usage, such observational data is nevertheless well suited for the development of theories that appear consistent with observations made by other authors (e.g., Palmer et al. 2017; Wassmann and Dasen 1998). We therefore argue that observations from fieldwork can motivate evolutionary hypotheses that can then be tested in the lab or modeled computationally (see Section 3).

A novel way of representing such theories in evolutionary linguistics is using "causal graphs" (Pearl 2009; Pearl and Mackenzie 2018), which allow hypotheses to be expressed and which take into account potential confounds.[2] A new tool, the Causal Hypotheses in Evolutionary Linguistics Database (CHIELD, https://chield.excd.org), is currently in development (see Roberts et al. 2020 for details on how to use the database). The idea behind CHIELD is that it allows causal claims extracted from research papers to be represented in the form of causal graphs, such as Figure 1, which shows the hypothesis graph based on a paper by Haviland (1993).

Haviland suggests that there is a causal link between the linguistic FoR and the gestural FoR, and more specifically that speakers preferring an absolute FoR relying on cardinal terms would orient their body according to cardinal directions as well when using co-speech gestures (e.g., when recalling a past event). All three links displayed in Figure 1 are based on qualitative observational data from video recordings of a single Guugu Yimithirr speaker. We can therefore consider them links that need to be empirically verified. CHIELD makes it easy to explore connections between documents through recurring causal links in the literature: The database reveals that another paper by Haun and Rapold (2009) did experimentally investigate the link between linguistic FoR and bodily orientation.[3] They tested how fifty German children and thirty-five children

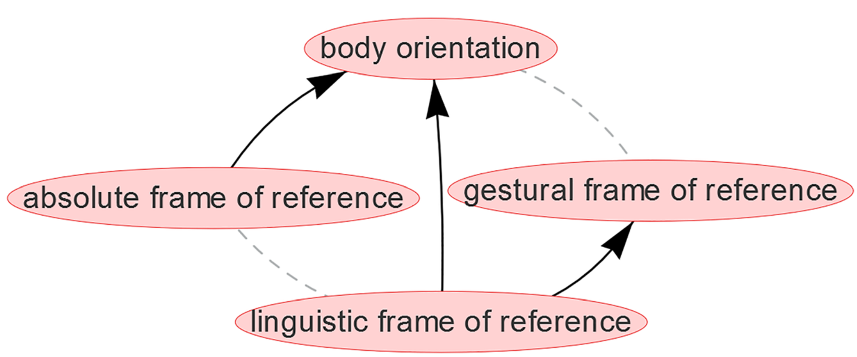

Figure 1: A simple causal graph based on hypothesized links between the use of an absolute frame of reference and body orientation during gesturing taken from Haviland (1993). The graph was generated using CHIELD (Roberts et al. 2020) and can be accessed at https://chield.excd.org/document.html?key=haviland1993anchoring#.

2 A detailed introduction to the construction of causal graphs is beyond the scope of this paper. For an overview, see McElreath 2020; Roberts et al. 2020: Chs. 5 and 6). A step-by-step guide for how to make hypotheses explicit and draw a causal graph is provided by Roberts (2018).

3 CHIELD allows specifying whether a link represents a hypothesis or is, e.g., based on qualitative or experimental evidence. Links can also be associated with a particular stage in language evolution (preadaptation, co-evolution, cultural evolution, or language change; see Scott-Phillips and Kirby 2010).

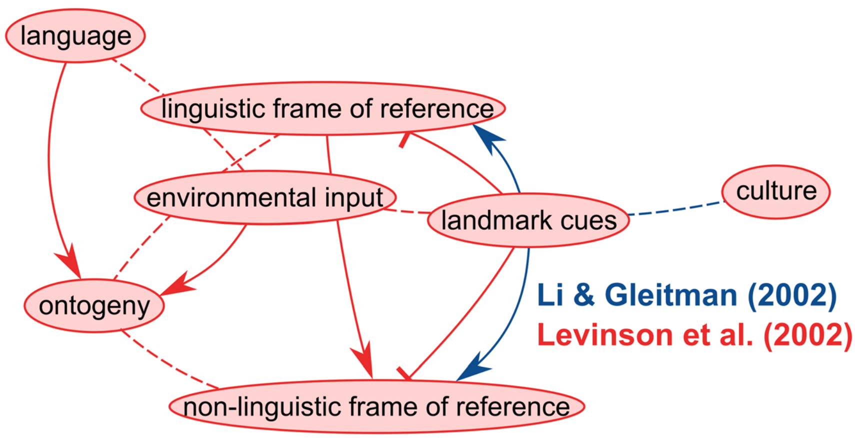

Figure 2: Example of a causal graph revealing a conflict in the literature. Note that X → Y specifies causal links, while X —| Y denotes that X has no causal influence on Y. Colors refer to specific publications. The graph was generated using CHIELD (Roberts et al. 2020).

speaking ≠Akhoe Hai||om – a language spoken in Namibia that prefers the absolute FoR – would orient their body when memorizing dance movements and found that their linguistic FoR preference significantly predicted their orientation during recall, supporting Haviland's (1993) original hypothesis.

In this way, causal graphs allow relationships between theorized variables to be represented and links that need to be empirically tested to be identified. Additionally, once research into spatial language has been made available in CHIELD, connections between studies can be used to identify conflicts; an example of this can be seen in Figure 2.

The causal graph in Figure 2 shows a well-known disagreement between Li and Gleitman (2002) and Levinson et al. (2002) on whether environmental landmark cues influence FoR choice on linguistic and cognitive tasks. As more papers are coded and causal graphs are added to the database, CHIELD can also represent more complex hypotheses involving many factors (topography, subsistence style, bilingualism, etc.). Formal representations of causal links are especially useful when more than three variables are involved (see, e.g., Roberts [2018] for a formalization of the relationship between climate and tone) and can even be derived algorithmically from observational data (Blasi and Roberts 2017).

It should be highlighted that observational data is essential in generating these hypothesis graphs, where each node represents a variable that has been observed to affect spatial language. For instance, work on Tseltal (Polian and Bohnemeyer 2011) suggests that local salient slopes could correlate with the use of a geocentric uphill–downhill FoR strategy (discussed in the following section). The next step is to convincingly demonstrate that this relationship can be attributed to a general mechanism that systematically shapes variation in spatial language based on the presence versus absence of particular types of landmarks. The following section outlines experimental and computational methods that can be used to model such mechanisms by studying the emergence and evolution of different spatial referencing systems under carefully controlled conditions.

3 Addressing causality with models and experiments

Beyond traditional cross-cultural experiments, evolutionary linguistics offers a methodological toolkit for addressing phenomena at various timescales, including modeling the ad hoc emergence of linguistic structure. While experiments have mostly focused on the basic building blocks of language like compositionality, combinatoriality, or arbitrariness (Tamariz 2017), recent studies address how external variables related to the environment or communicative context can motivate systematic differences between artificial languages emerging in interaction (Winters et al. 2015) or over cultural transmission (Tinits et al. 2017). For instance, Raviv et al. (2019) found that artificial languages emerging in larger laboratory communities develop a more systematic grammar, which supports earlier statistical observations suggesting relations between linguistic and social structure (Lupyan and Dale 2010). Another study by Nölle et al. (2020a) tested whether environmental affordances could affect two-dimensional spatial conceptualizations in the Maze Game, a collaborative task that has previously been used to study the emergence and diffusion of linguistic conventions in dialog (Garrod and Doherty 1994). They found that the presence of specific maze configurations and salient landmarks (e.g., figural shapes) predicted how participants would describe locations in the maze, suggesting that spatial linguistic strategies are sensitive to salient landmarks in the environment where they are grounded.

A recent innovation of laboratory paradigms is the possibility of studying spatial language on a "large" scale. Classical tasks used to study spatial referencing and cognition cross-culturally, such as Man and Tree (Levinson et al. 1992), Ball and Chair (e.g., O'Meara and Pérez Báez 2011), and the Animals in a Row task (Levinson and Schmitt 1993), usually involve relations between objects in small-scale tabletop space or photographs. While these tasks are very effective in evoking spatial linguistic descriptions, these descriptions can be highly task-specific and are not necessarily informative about the use of FoR with respect to distant objects or landmarks. In the real world, geocentric spatial language is usually embedded in situated social interactions involving wayfinding or other spatial tasks in complex large-scale environments: For instance, Palmer and colleagues report that Dhivehi speakers on fishing islands use more geocentric descriptions than speakers on non-fishing islands, and similarly that Marshallese speakers living in Springdale Arkansas, who have an urban lifestyle and participate in activities such as driving, use more egocentric and less geocentric descriptions than members of speech communities on the Marshall Islands (Palmer et al. 2017, 2018). How can potential links between these variables be tested? It appears difficult to support such statistical findings with controlled laboratory experiments that try to isolate the causal contribution of factors such as engagement with the sea or urban mobility. Real-world environments are complex, noisy, and hard to control across many experimental trials. Traditionally, experimental control and ecological validity have therefore been considered the ends of a continuum (Peeters 2019): at one end, approaches like conversation analysis allow observing communication in its full multimodal richness, but necessarily sacrifice control over participants' behavior and communicative conditions; while at the other end, highly abstract tasks, like picture-naming or reaction time experiments, allow researchers to isolate specific conditions and phenomena but are far removed from naturalistic conversation.

However, as suggested by Peeters (2019), it is increasingly becoming clear that experimental control and ecological validity are two orthogonal factors of experimental design, rather than the ends of a continuum. It is now possible to generate finely controlled virtual environments that remain constant over experimental sessions and trials. For instance, Lum and Schlossberg (2014) devised the Virtual Atoll Task (VAT), a virtual large-scale analog of classical director-matcher tasks, such as the Man and Tree game, where participants explore a virtual 3D model of an atoll displayed on a computer screen. The task was meant to elicit large-scale spatial descriptions in a more realistic setting closer to real-world wayfinding. While some limitations resulted from the task design and controls (which could bias egocentric behavior), the VAT was the first experiment to successfully demonstrate the elicitation of spatial language in a more naturalistic large-scale environment.

Immersive virtual reality (VR) paradigms that enable even more realistic interaction with a virtual world are already established in rehabilitation therapy (Bohil et al. 2011; Gould et al. 2007) and spatial navigation research (Tarr and Warren 2002), but have only recently been introduced to the language sciences (Heyselaar et al. 2017; Peeters 2019). VR setups have great potential for studying spatial language, since they allow researchers to control factors such as the environmental layout, the size and shape of objects, and participants' orientation and perspective, and to model field sites that are normally not easily accessible.

Additionally, VR allows for the manipulation of otherwise realistic tasks in unprecedented ways. For example, Nölle et al. (2020b) describe OrbHunt, a collaborative VR game where a "director" has to communicate the location of "orbs" to a "seeker" in order to score points. This is very similar to the Maze Game (Garrod and Doherty 1994; Nölle et al. 2020a) in that location descriptions are necessary to solve the task. However, in OrbHunt both players have a natural first-person perspective and elicited descriptions refer to locations in actual 3D space. VR further allows researchers to create an incentive for communicating spatial locations by limiting the distance from which the seeker can see orbs (which cannot be controlled in the real world), and it allows for a comparison of the exact same task across different topographic environments, namely a noticeably sloping environment versus a flat environment that resembled a dense forest (see Figure 3). Although both

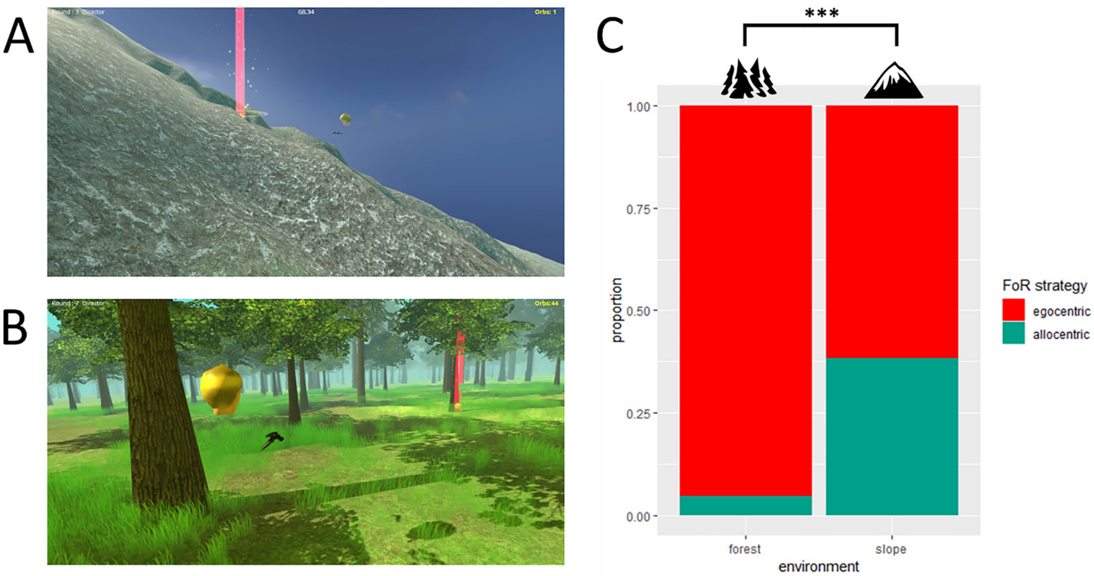

Figure 3: Screenshots from the OrbHunt game (director's viewpoint) in the slope (A) and forest (B) environments. English-speaking participants were significantly less likely to rely purely on egocentric descriptions if they were playing on the mountain slope (C).

environments afforded fully egocentric solutions and participants were English speakers, who generally prefer an egocentric FoR, it was found that FoR strategies differed depending on whether a salient uphill–downhill axis was present, which afforded geocentric descriptions similar to Tseltal (Polian and Bohnemeyer 2011). While dyads used a multitude of description strategies, they relied less on the egocentric FoR when playing on the slope and utilized more allocentric strategies, that is, strategies based on geocentric *uphill*, *downhill*, and *across* (see Figure 3C). This provides experimental evidence for the hypothesis that the presence of landmark cues can motivate specific linguistic FoRs (compare Figure 2).

Nölle et al. (2020b) also undertook a second experiment, which tested how spatial referencing strategies adapt to shifting environments. In two conditions, dyads either started playing OrbHunt in the forest or on the slope and switched to the other environment after five rounds. Results indicated that participants in both conditions (and thus block orders, i.e., sequences in which the environments were encountered) were readily able to flexibly switch their FoR strategies in response to the environment at hand (Figure 4).

By isolating the environment as an independent variable, these VR experiments demonstrate that, in ecologically rich settings, topographic cues alone can affect the spatial referencing strategies chosen by English speakers to communicate locations on a large scale. Future studies can extend these findings and compare speakers of different languages (e.g., those relying on a predominantly geocentric system) and different environments (e.g., urban vs. rural), as well as different tasks and scales. The advantage of VR really lies in the fact that all aspects of the task can be tightly controlled. For instance, it is possible to test whether participants adapt the strategy they use to reference configurations of objects on the large scale when they have to suddenly switch to talking about arrays of the same objects shrunk to tabletop space. This could help address questions about how different scales and the existence of manipulable objects affect FoR use (see Bohnemeyer et al. 2018).

An additional OrbHunt experiment showed that it is also possible to study the emergence of spatial referencing systems per se and how such systems evolve in response to external conditions (see Nölle 2021: Ch. 5). The experiment took an "experimental semiotics" approach (Galantucci et al. 2012), where participants solving the OrbHunt game were deprived of their ability to use language and had to invent a novel communication system via visual signals instead. It was found that participants were able to do so successfully, and that movement patterns and gaze directions recorded in the VR could be used to draw conclusions about which FoR strategies were associated with the signals they sent to each other. This study is thus a first preliminary step for modeling

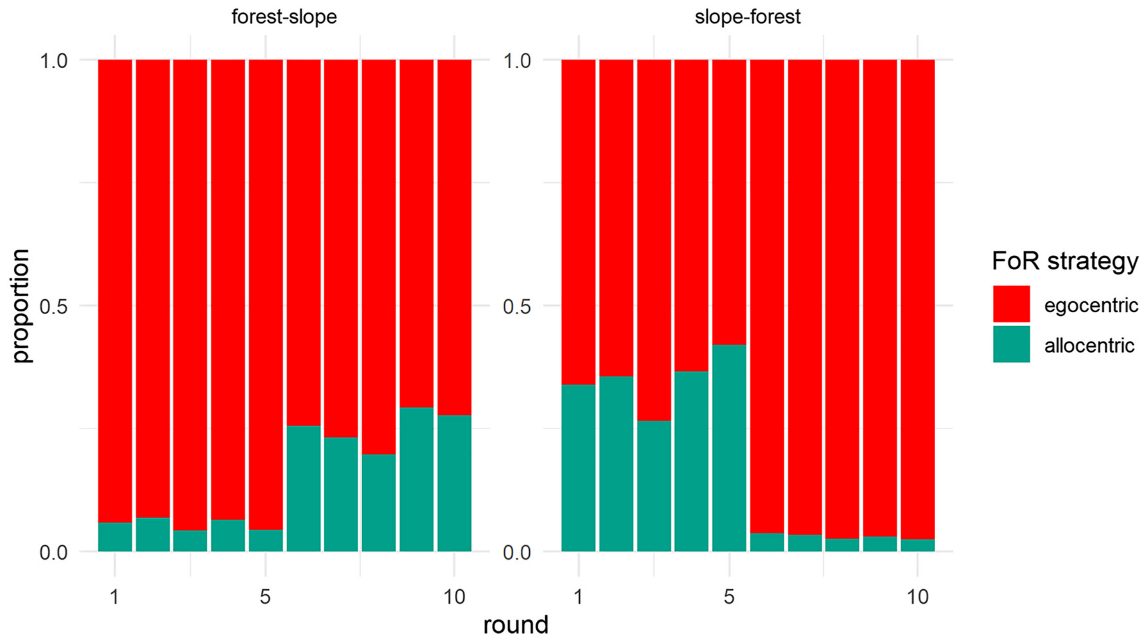

Figure 4: Results for the second OrbHunt experiment (Nölle et al. 2020b) showing proportions of FoR strategies over ten rounds for block orders. Participants started in either the forest (left) or the slope (right) and switched environments after playing five rounds.

the possible stages in the evolution of spatial referencing systems and how these systems are affected by internal cognitive biases and sociotopographic variables, which can then be compared against real-world data to get a mechanistic picture of how spatial linguistic diversity is motivated and transmitted culturally.

Finally, such laboratory experiments can be integrated with agent-based models that simulate how spatial language might evolve in the context of various factors: environmental constraints, sensorimotor perception, cognitive abilities, social interaction, and cultural dynamics (Spranger 2016). The advantage of such models is that they can simulate the evolution of spatial referencing systems over larger timescales and bigger populations, and systematically explore how small changes in the environment (number of objects, availability of geocentric landmarks, etc.) or the agents' cognitive abilities affect the evolution of FoR systems. Importantly, these experiments complement behavioral experiments by explicitly and transparently modeling agent-internal mechanisms.

For instance, robotic models by Spranger (2011, 2013) show (1) how spatial conceptualization strategies can be represented in the memory of each agent and (2) how social interactions between agents of a population give rise to emerging language conventions. In particular, Spranger shows how spatial conceptualization strategies can be built by individual agents in a process of recruitment of general cognitive abilities such as categorization and perspective reversal, and how these strategies can become conventionalized through repeated interactions of agents of a population. The work shows that particular FoR choices of a population can arise in agents with the same cognitive abilities but in different ecological conditions. For example, Figure 5 shows the results of different populations evolving spatial language in environments with different availability of geocentric (environment-centered) or intrinsic (object-centered) features and the effect of this on the evolution of the communication systems.

Other agent-based models have focused on the emergence of spatial categories (Spranger 2012a), landmark-based strategies (Spranger 2012b), and spatial grammar (Spranger and Steels 2012). The research primarily shows that agents need internal representations of competing conventions for spatial conceptualization and their expression. The studies also identify specific mechanisms for the agent-internal update of these strategies to achieve population-level alignment and how population choices are influenced by ecological conditions (Spranger 2016).

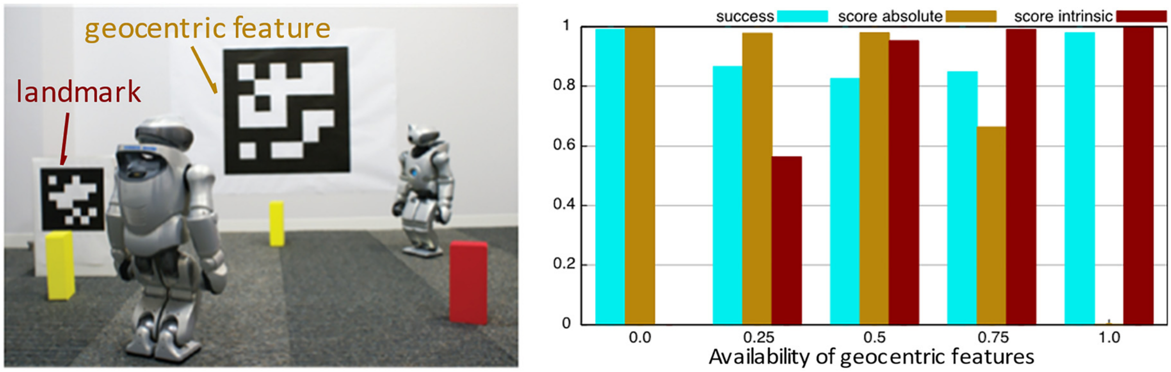

Figure 5: Spatial linguistic systems created by robots in interaction reflect the presence and absence of intrinsic and absolute (i.e., geocentric) features in their environment. Left panel: Robotic agents have to solve referential communication games communicating about colored objects in the scene. To disambiguate referents, they can rely on intrinsic properties of objects (e.g., *the yellow block in front of the landmark*) or invariant geocentric features (e.g., *the yellow block north of the red block*). Right panel: Absolute spatial referencing systems evolve in environments where only geocentric features are available (far left on the *x*-axis), while intrinsic systems develop when only intrinsic features are available (far right on the *x*-axis). If both types of features are present, the agents tend to develop mixed systems that rely on both FoRs. See Spranger (2016) for further details.

4 Conclusions

Languages are complex adaptive systems that evolve non-linearly according to principles of self-organization and selection in interaction with internal and external constraints (Beckner et al. 2009). Recent typological work (reviewed in Section 2) has revealed that, similarly, variation in spatial referencing presents a complex

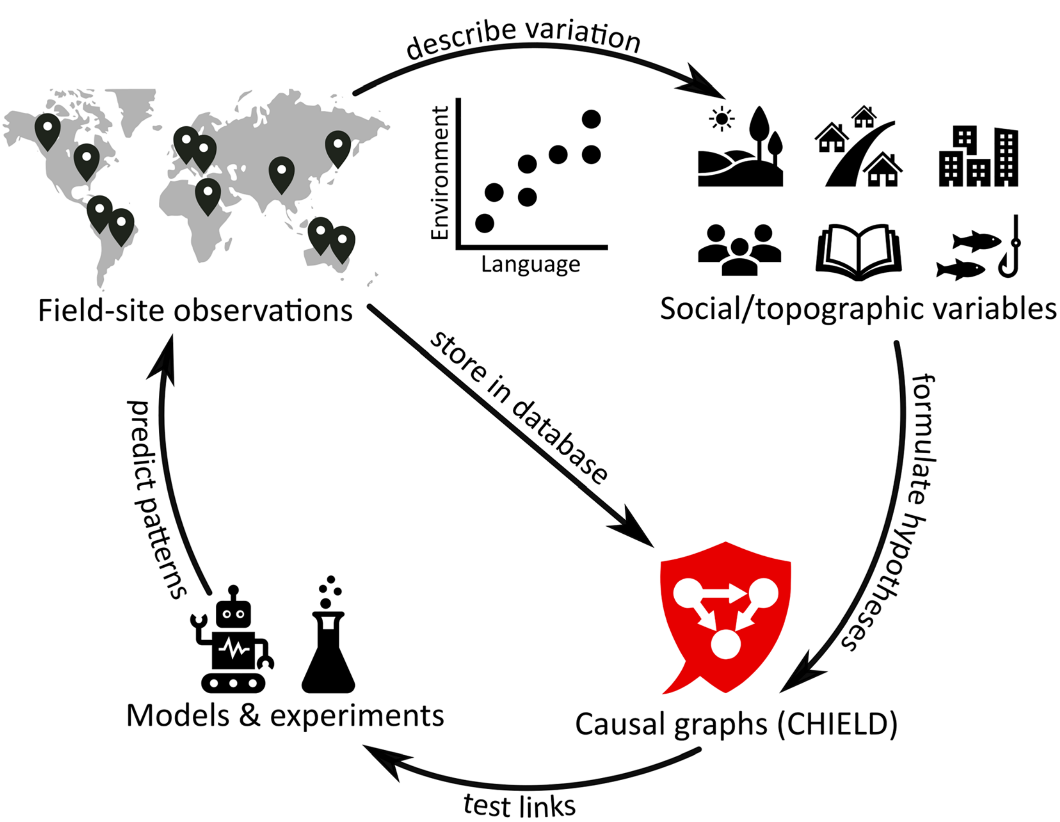

Figure 6: Schematic cycle using the maximum robustness approach to investigate the relationship between spatial language and sociotopographic variables.

problem at the intersection of language, culture, and cognition that can only be addressed with multiple methods in order to disentangle how different variables interact with FoR choice.

In this paper, we have suggested that an interdisciplinary "maximum robustness" approach (Roberts 2018) in combination with experiments and computational models simulating the origins and evolution of spatial language can fill this gap. Initially, descriptive, cross-sectional work is necessary to identify relevant patterns in the form of statistical tendencies among speech communities. We have shown how causal graphs can be used to explicitly define hypotheses, which can then be incrementally tested to reveal robust relationships using experiments that can test causal links, avoid spurious correlations and model the emergence of variation on a more mechanistic level (see Figure 6).

In this regard, research into spatial language can benefit from novel methodologies that have recently been introduced to study the cultural evolution of language. Computational models can simulate the evolution and transmission of spatial language in large populations over long timescales while accounting for internal and external mechanisms. VR experiments with human participants can be used to study actual human behavior under controlled conditions with increased ecological validity. Future studies relying on a combination of these methods can address a variety of questions: Are egocentric systems more flexible than geocentric ones? Are egocentric FoRs more easily applied to the small scale (i.e., tabletop space)? Which systems fare better in different urban environments? What activities favor the integration of geocentric cues? How do strategies diffuse within and across communities? Taken together, this evolutionary approach can address questions derived from qualitative observations by explicitly modeling the potential stages in the cultural evolution of spatial language and shed light on how sociotopographic factors shape the variation in spatial referencing systems among the world's languages.

Acknowledgments: This paper benefited from discussions at the specialist meeting "Universals and Variation in Spatial Referencing across Cultures and Languages" hosted at the Center for Spatial Studies at UC Santa Barbara in December 2016. In particular, we wish to thank Bill Palmer, Jürgen Bohnemeyer, Niclas Burenhult, Terry Regier, and Steve Levinson for insightful discussions. We also wish to thank Alice Gaby, Stefan Hartmann, and two anonymous reviewers for comments on earlier versions of this article.

References

Adamou, Evangelia & Xingjia Rachel Shen. 2017. Beyond language shift: Spatial cognition among the Ixcatecs in Mexico. *Journal of Cognition and Culture* 17(1–2). 94–115.

Ameka, Felix K. & James Essegbey. 2006. Elements of the grammar of space in Ewe. In Stephen C. Levinson & David P. Wilkins (eds.), *Grammars of space: Explorations in cognitive diversity*, 359–399. Cambridge: Cambridge University Press.

Beckner, Clay, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman & Tom Schoenemann. 2009. Language is a complex adaptive system: Position paper. *Language Learning* 59. 1–26.

Bentz, Christian & Bodo Winter. 2013. Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change* 3(1). 1–27.

Blasi, Damián E., Steven Moran, Scott R. Moisik, Paul Widmer, Dan Dediu & Balthasar Bickel. 2019. Human sound systems are shaped by post-Neolithic changes in bite configuration. *Science* 363(6432). eaav3218.

Blasi, Damián E. & Seán G. Roberts. 2017. Beyond binary dependencies in language structure. In Nick J. Enfield (ed.), *Dependencies in language: On the causal ontology of linguistic systems*, 117–128. Berlin: Language Science Press.

Bohil, Corey J., Bradly Alicea & Frank A. Biocca. 2011. Virtual reality in neuroscience research and therapy. *Nature Reviews Neuroscience* 12(12). 752–762.

Bohnemeyer, Jürgen. 2011. Spatial frames of reference in Yucatec: Referential promiscuity and task-specificity. *Language Sciences* 33(6). 892–914.

Bohnemeyer, Jürgen. 2016. The interplay of language, culture, and environment in spatial reference. Paper presented at the conference Universals and variation in spatial referencing across cultures and languages. Santa Barbara: Center for Spatial Studies.

Bohnemeyer, Jürgen, Katharine Donelson, Yen-Ting Lin & Randi Moore. 2018. The biological and cultural evolution of spatial cognition: Evidence from semantic typology. In Thomas Hünefeldt & Marta Olivetti Belardinelli (eds.), *Spatial Cognition in a* *multimedia and intercultural world. Proceedings of the 7th International Conference on Spatial Cognition (ICSC 2018)*. Rome: Springer.

Bohnemeyer, Jürgen, Katharine T. Donelson, Randi E. Moore, Elena Benedicto, Alyson Eggleston, Carolyn K. O'Meara, Gabriela Pérez Báez, Alejandra Capistrán Garza, Néstor Hernández Green & María de Jesús Selene Hernández Gómez. 2015. The contact diffusion of linguistic practices. *Language Dynamics and Change* 5(2). 169–201.

Brown, Angela M. & Delwin T. Lindsey. 2004. Color and language: Worldwide distribution of Daltonism and distinct words for "blue". *Visual Neuroscience* 21(3). 409–412.

Calderón, Eréndira, Stefano De Pascale & Evangelia Adamou. 2019. How to speak "geocentric" in an "egocentric" language: A multimodal study among Ngigua-Spanish bilinguals and Spanish monolinguals in a rural community of Mexico. *Language Sciences* 74. 24–46.

Chen, M. Keith. 2013. The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. *The American Economic Review* 103(2). 690–731.

Coupé, Christophe, Yoon Oh, Dan Dediu & François Pellegrino. 2019. Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Science Advances* 5(9). eaaw2594.

Danziger, Eve & Eric Pederson. 1998. Through the looking glass: Literacy, writing systems and mirror-image discrimination. *Written Language & Literacy* 1(2). 153–169.

Dasen, Pierre R. & Ramesh C. Mishra. 2010. *Development of geocentric spatial language and cognition: An eco-cultural perspective*. Cambridge: Cambridge University Press.

Dediu, Dan, Michael Cysouw, Stephen C. Levinson, Andrea Baronchelli, Morten H. Christiansen, William Croft, Nicholas Evans, Simon Garrod, Russell D. Gray, Anne Kandler & Elena Lieven. 2013. Cultural evolution of language. In Peter J. Richerson & Morten H. Christiansen (eds.), *Cultural evolution: Society, technology, language, and religion*, 303–332. Cambridge, MA: MIT Press.

Dediu, Dan, Rick Janssen & Scott R. Moisik. 2017. Language is not isolated from its wider environment: Vocal tract influences on the evolution of speech and language. *Language & Communication* 54. 9–20.

Diessel, Holger. 2014. Demonstratives, frames of reference, and semantic universals of space. *Linguistics and Language Compass* 8(3). 116–132.

Donelson, Katharine. 2018. *On space in conversation: Coordinating perspectives in spatial reference*. Buffalo, NY: State University of New York at Buffalo PhD thesis.

Eggleston, Alyson, Elena Benedicto & Mayangna Yulbarangyang Balna. 2011. Spatial frames of reference in Sumu-Mayangna. *Language Sciences* 33(6). 1047–1072.

Everett, Caleb, Damián E. Blasi & Seán G. Roberts. 2015. Climate, vocal folds, and tonal languages: Connecting the physiological and geographic dots. *Proceedings of the National Academy of Sciences* 112(5). 1322–1327.

Feldmann, Horst. 2019. Do linguistic structures affect human capital? The case of pronoun drop: Linguistic structures and human capital. *Kyklos* 72(1). 29–54.

Galantucci, Bruno, Simon Garrod & Gareth Roberts. 2012. Experimental semiotics. *Language and Linguistics Compass* 6(8). 477–493.

Gallistel, Randy C. 2002. Conception, perception and the control of action. *Trends in Cognitive Sciences* 6(12). 504.

Garrod, Simon & Gwyneth Doherty. 1994. Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition* 53(3). 181–215.

Gould, Neda F., M. Kathleen Holmes, Bryan D. Fantie, David A. Luckenbaugh, Daniel S. Pine, Todd D. Gould, Neil Burgess, Husseini K. Manji & Carlos A. Zarate. 2007. Performance on a virtual reality spatial memory navigation task in depressed patients. *American Journal of Psychiatry* 4. 516–519.

Haun, Daniel B. M. & Christian J. Rapold. 2009. Variation in memory for body movements across cultures. *Current Biology* 19(23). R1068–R1069.

Haun, Daniel B. M., Christian J. Rapold, Josep Call, Gabriele Janzen & Stephen C. Levinson. 2006. Cognitive cladistics and cultural override in Hominid spatial cognition. *Proceedings of the National Academy of Sciences* 103(46). 17568–17573.

Haun, Daniel B. M., Christian J. Rapold, Gabriele Janzen & Stephen C. Levinson. 2011. Plasticity of human spatial cognition: Spatial language and cognition covary across cultures. *Cognition* 119(1). 70–80.

Haviland, John B. 1993. Anchoring, iconicity, and orientation in Guugu Yimithirr pointing gestures. *Journal of Linguistic Anthropology* 3(1). 3–45.

Heyselaar, Evelien, Peter Hagoort & Katrien Segaert. 2017. In dialogue with an avatar, language behavior is identical to dialogue with a human partner. *Behavior Research Methods* 49(1). 46–60.

Kim, Jaehyeon, Yongtae Kim & Jian Zhou. 2017. Languages and earnings management. https://papers.ssrn.com/abstract=2941120 (accessed 6 August 2019).

Lawton, Carol A. 2001. Gender and regional differences in spatial referents used in direction giving. *Sex Roles* 44(5–6). 321–337.

Le Guen, Olivier. 2011. Speech and gesture in spatial language and cognition among the Yucatec Mayas. *Cognitive Science* 35(5). 905–938.

Levinson, Stephen C. 2003. *Space in language and cognition: Explorations in cognitive diversity*. Cambridge: Cambridge University Press.

Levinson, Stephen C., Penelope Brown, Eve Danziger, Lourdes De Léon, John B. Haviland, Eric Pederson & Gunter Senft. 1992. Man and tree & space games. In Stephen C. Levinson (ed.), *Space stimuli kit 1.2: November 1992*, 7–14. Nijmegen: Max Planck Institute for Psycholinguistics.
Levinson, Stephen C., Sotaro Kita, Daniel B. M. Haun & Björn H. Rasch. 2002. Returning the tables: Language affects spatial reasoning. *Cognition* 84(2). 155–188.
Levinson, Stephen C. & Bernadette Schmitt. 1993. Animals in a row. In *Cognition and space kit version 1.0.*, 65–69. Nijmegen: Cognitive Anthropology Research Group at the Max Planck Institute for Psycholinguistics.
Levinson, Stephen C. & David P. Wilkins. 2006. *Grammars of space: Explorations in cognitive diversity*. Cambridge: Cambridge University Press.
Li, Peggy, Linda Abarbanell, Lila Gleitman & Anna Papafragou. 2011. Spatial reasoning in Tenejapan Mayans. *Cognition* 120(1). 33–53.
Li, Peggy & Lila Gleitman. 2002. Turning the tables: Language and spatial reasoning. *Cognition* 83(3). 265–294.
Lin, Yen-Ting. 2017. *Spatial language and cognition in bilingual minds: Taiwan as a test case*. Buffalo, NY: State University of New York at Buffalo PhD thesis.
Lum, Jonathon & Jonathan Schlossberg. 2014. The Virtual Atoll Task: A spatial language elicitation tool. In Mark Harvey & Alexis Antonia (eds.), *The 45th Australian Linguistic Society Conference Proceedings – 2014*. Newcastle: NOVA Open Access Repository. Available at: http://hdl.handle.net/1959.13/1307976.
Lupyan, Gary & Rick Dale. 2010. Language structure is partly determined by social structure. *PLOS ONE* 5(1). e8559.
Lupyan, Gary & Rick Dale. 2016. Why are there different languages? The role of adaptation in linguistic diversity. *Trends in Cognitive Sciences* 20(9). 649–660.
Majid, Asifa, Melissa Bowerman, Sotaro Kita, Daniel B. M. Haun & Stephen C. Levinson. 2004. Can language restructure cognition? The case for space. *Trends in Cognitive Sciences* 8(3). 108–114.
McElreath, Richard. 2020. *Statistical rethinking: A Bayesian course with examples in R and Stan*, 2nd edn. Boca Raton, FL: CRC Press.
Mesoudi, Alex & Alex Thornton. 2018. What is cumulative cultural evolution? *Proceedings of the Royal Society B: Biological Sciences* 285(1880). 20180712.
Mishra, Ramesh C., Pierre R. Dasen & Shanta Niraula. 2003. Ecology, language, and performance on spatial cognitive tasks. *International Journal of Psychology* 38(6). 366–383.
Moore, Randi. 2018. *Spatial language and cognition in Isthmus Zapotec*. Buffalo, NY: State University of New York at Buffalo PhD thesis.
Naroll, Raoul. 1961. Two solutions to Galton's problem. *Philosophy of Science* 28(1). 15–39.
Newcombe, Nora S. & Janellen Huttenlocher. 2000. *Making space: The development of spatial representation and reasoning*. Cambridge, MA: MIT Press.
Nölle, Jonas. 2021. *How language adapts to the environment: An evolutionary, experimental approach*. Edinburgh: University of Edinburgh PhD thesis.
Nölle, Jonas, Riccardo Fusaroli, Gregory J. Mills & Kristian Tylén. 2020a. Language as shaped by the environment: Linguistic construal in a collaborative spatial task. *Palgrave Communications* 6(1). 1–10.
Nölle, Jonas, Simon Kirby, Jennifer Culbertson & Kenny Smith. 2020b. Does environment shape spatial language? A virtual reality approach. In Andrea Ravignani, Chiara Barbieri, Mauricio Martins, Molly Flaherty, Yannick Jadoul, Ella Lattenkamp, Hannah Little, Katie Mudd & Tessa Verhoef (eds.), *The evolution of language: Proceedings of the 13th international conference (Evolang XIII)*, 321–323. Nijmegen: Max Planck Institute for Psycholinguistics.
O'Meara, Carolyn & Gabriela Pérez Báez. 2011. Spatial frames of reference in Mesoamerican languages. *Language Sciences* 33(6). 837–852.
Palmer, Bill. 2015. Topography in language: Absolute Frame of Reference and the Topographic Correspondence Hypothesis. In Rik De Busser & Randy J. LaPolla (eds.), *Language structure and environment: Social, cultural, and natural factors*, 177–226. Amsterdam: John Benjamins.
Palmer, Bill, Alice Gaby, Jonathan Lum & Jonathan Schlossberg. 2018. Diversity in spatial language within communities: The interplay of culture, language and landscape in representations of space (short paper). In *10th International Conference on Geographic Information Science (GIScience 2018) 2018. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik*, 1–8. https://doi.org/10.4230/lipics.giscience.2018.53.
Palmer, Bill, Jonathon Lum, Jonathan Schlossberg & Alice Gaby. 2017. How does the environment shape spatial language? Evidence for sociotopography. *Linguistic Typology* 21(3). 457–491.
Pearl, Judea. 2009. *Causality*. Cambridge: Cambridge University Press.
Pearl, Judea & Dana Mackenzie. 2018. *The book of why: The new science of cause and effect*. USA: Basic Books.
Pederson, Eric, Eve Danziger, David Wilkins, Stephen C. Levinson, Sotaro Kita & Gunter Senft. 1998. Semantic typology and spatial conceptualization. *Language* 74(3). 557–589.
Peeters, David. 2019. Virtual reality: A game-changing method for the language sciences. *Psychonomic Bulletin & Review* 26. 894–900.
Pinker, Steven. 2007. *The stuff of thought: Language as a window into human nature*. London: Penguin.

Polian, Gilles & Jürgen Bohnemeyer. 2011. Uniformity and variation in Tseltal reference frame use. *Language Sciences* 33(6). 868–891.
Raviv, Limor, Antje Meyer & Shiri Lev-Ari. 2019. Larger communities create more systematic languages. *Proceedings of the Royal Society B: Biological Sciences* 286(1907). 20191262.
Regier, Terry, Alexandra Carstensen & Charles Kemp. 2016. Languages support efficient communication about the environment: Words for snow revisited. *PLOS ONE* 11(4). e0151138.
Richerson, Peter J. & Robert Boyd. 2008. *Not by genes alone: How culture transformed human evolution*. Chicago: University of Chicago Press.
Roberts, Seán G. 2018. Robust, causal, and incremental approaches to investigating linguistic adaptation. *Frontiers in Psychology* 9. 1–21.
Roberts, Seán G., Anton Killin, Angarika Deb, Catherine Sheard, Simon J. Greenhill, Kaius Sinnemäki, José Segovia-Martín, Jonas Nölle, Aleksandrs Berdicevskis, Archie Humphreys-Balkwill, Hannah Little, Christopher Opie, Guillaume Jacques, Lindell Bromham, Peeter Tinits, Robert M. Ross, Sean Lee, Emily Gasser, Jasmine Calladine, Matthew Spike, Stephen F. Mann, Olena Shcherbakova, Ruth Singer, Shuya Zhang, Antonio Benítez-Burraco, Christian Kliesch, Ewan Thomas-Colquhoun, Hedvig Skirgård, Monica Tamariz, Monica Tamariz, Sam Passmore, Thomas Pellard & Fiona Jordan. 2020. CHIELD: The causal hypotheses in evolutionary linguistics database. *Journal of Language Evolution* 5(1). 101–120.
Roberts, Seán G. & James Winters. 2013. Linguistic diversity and traffic accidents: Lessons from statistical studies of cultural traits. *PLOS ONE* 8(8). e70902.
Scott-Phillips, Thomas C. & Simon Kirby. 2010. Language evolution in the laboratory. *Trends in Cognitive Sciences* 14(9). 411–417.
Shapero, Joshua A. 2017. Does environmental experience shape spatial cognition? Frames of reference among Ancash Quechua speakers (Peru). *Cognitive Science* 41(5). 1274–1298.
Shusterman, Anna & Peggy Li. 2016. Frames of reference in spatial language acquisition. *Cognitive Psychology* 88. 115–161.
Spranger, Michael. 2011. Recruitment, selection and alignment of spatial language strategies. In Tom Lenaerts, Mario Giacobini, Hugues Bersini, Paul Bourgine, Marco Dorigo & René Doursat (eds.), *Advances in artificial life, ECAL 2011: Proceedings of the eleventh European Conference on the Synthesis and Simulation of Living Systems*, 771–778. Cambridge, MA: MIT Press.
Spranger, Michael. 2012a. The co-evolution of basic spatial terms and categories. In Luc Steels (ed.), *Experiments in cultural language evolution*, 111–141. Amsterdam: John Benjamins.
Spranger, Michael. 2012b. Potential stages in the cultural evolution of spatial language. In Thomas C. Scott-Phillips, Mónica Tamariz, Erica A. Cartmill & James R. Hurford (eds.), *The evolution of language: Proceedings of the 9th conference (Evolang IX)*, 328–335. Kyoto: World Scientific.
Spranger, Michael. 2013. Evolving grounded spatial language strategies. *KI – Künstliche Intelligenz* 27(2). 97–106.
Spranger, Michael. 2016. *The evolution of grounded spatial language*. Berlin: Language Science Press.
Spranger, Michael & Luc Steels. 2012. Synthetic modeling of cultural language evolution. In Luke McCrohon, Tomomi Fujimura, Kazuo Okanoya, Koji Fujita, Reiji Suzuki, Roger Martin & Noriaki Yusa (eds.), *Five approaches to language evolution – Proceedings of the workshops of the 9th International Conference on the Evolution of Language (Evolang IX)*, 130–139. Kyoto: World Scientific.
Tamariz, Monica. 2017. Experimental studies on the cultural evolution of language. *Annual Review of Linguistics* 3(1). 389–407.
Tarr, Michael J. & William H. Warren. 2002. Virtual reality in behavioral neuroscience and beyond. *Nature Neuroscience* 5(11). 1089–1092.
Tinits, Peeter, Jonas Nölle & Stefan Hartmann. 2017. Usage context influences the evolution of overspecification in iterated learning. *Journal of Language Evolution* 2(2). 148–159.
Wassmann, Jurg & Pierre R. Dasen. 1998. Balinese spatial orientation: Some empirical evidence of moderate linguistic relativity. *The Journal of the Royal Anthropological Institute* 4(4). 689–711.
Winters, James, Simon Kirby & Kenny Smith. 2015. Languages adapt to their contextual niche. *Language and Cognition* 7(3). 415–449.