



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

How Language Adapts to the Environment: An Evolutionary, Experimental Approach

Jonas Nölle

2021

Submitted in fulfilment of the degree of Doctor of Philosophy to
School of Philosophy, Psychology and Language Sciences
University of Edinburgh

To my parents and grandparents

Lay Summary

Why do humans speak different languages instead of the same? Is this simply the outcome of arbitrary cultural conventions that we decided upon, or can linguistic diversity be explained by other factors? This thesis explores whether human languages adapt to their environment in a similar way to organisms in biological evolution. To test whether this is the case, I conducted two series of experiments. In the first series pairs of participants came into the lab and had to play a game where they could only use their hands to communicate about cartoon pictures. In two experiments we found that the way in which they systematically picked up traits from these cartoon pictures, and which traits they chose, depended on the context, for example how many targets they were communicating about, and whether they could see the targets or had to recall them from memory. This suggests that a fundamental property of human languages, the ability to mark systematic categories depends, to some extent, on the environment in which these languages develop, which could lead to diversity. A second series of experiments made use of virtual reality games, where participants had to communicate about the location of orbs. Different languages have different preferences for how they express spatial relations, e.g., *the orb is to the left of you* vs *the orb is uphill from you*. The former strategy can be called egocentric because it depends on an observer's viewpoint, while the latter strategy can be called geocentric because the orb is located with respect to an environmental property. By having participants play the same game either in a virtual forest or on a virtual mountain slope, I found that this environmental manipulation could systematically motivate these spatial strategies, with participants using more egocentric language in the forest, and more geocentric language on the slope. I also found that if participants switch environments, they can readily adapt their spatial language if they are familiar with these different ways to talk about space. However, when they could not use their native language and had to communicate via visual symbols, they all defaulted to the egocentric strategy. Taken together, this thesis presents novel experimental results suggesting that indeed, some variation between languages appears to result from adaptation to environments. Furthermore, the thesis makes various methodological contributions to the field of language evolution research by introducing virtual reality experiments and novel analysis techniques.

Abstract

The aim of this thesis is to investigate experimentally whether cross-linguistic variation in the structure of languages can be motivated by their external environment. It has been suggested that variation does not only result from cultural drift and language-internal mechanisms but also from social or even physical factors. However, from observational data and correlations between variables alone, it remains difficult to infer the exact underlying mechanisms.

Here, I present a novel experimental approach for studying the relationship between language and environment under controlled laboratory conditions. I argue that to arrive at a causal understanding of linguistic adaptation, we can use a cultural evolutionary approach and simulate the emergence of linguistic structure with humans in the lab. This way, it can be tested which pressures shape linguistic features as they are used for communication and transmitted to new speakers. I focus primarily on cases where linguistic conventions emerge in referential communication games in direct face-to-face interaction. In these settings, I test whether specific conventions are more adaptive to solve the same problem under different conditions or affordances imposed by the environment. A series of silent-gesture experiments shows that systematicity (the design feature giving language its compositional power) is sensitive to the communicative environment: Dyads creating novel gestural communication systems to communicate pictorial referents are more likely to systematize traits and create categories that are functionally relevant in the given environment. Additionally, environmental features, such as the size of the meaning space and visibility of referents, affect the degree to which participants rely on systematic rather than simple holistic gestures. This ‘experimental semiotics’ approach thus models how environmental factors could motivate basic linguistic structure.

However, for complex real-world phenomena, such as the hotly debated relationship between spatial language and environment, it is difficult to design simple experiments that isolate variables of interest but retain the necessary level of realism. It has been proposed that topography (e.g., landmarks like rivers, slopes) and sociocultural factors (e.g., bilingualism, subsistence style, population density) can affect whether speakers rely on an egocentric or geocentric Frame of Reference (FoR) to encode spatial relations, but it remains hard to disentangle the exact contribution of these variables to the cross-linguistic variation we observe.

I tackle this issue with a novel paradigm: interactive Virtual Reality (VR) experiments that allow for an unprecedented combination of ecological validity and experimental control. In networked VR settings, participants are immersed in realistic settings such as a forest or a mountain slope. By having dyads solve

spatial coordination games, I show that speakers of English, which is usually associated with an egocentric FoR, are less likely to use egocentric language (e.g., “the orb is to your left”) if there are strong environmental affordances that make geocentric language more viable (e.g., “the orb is uphill from you”). Further experiments address whether the cultural ‘success’ of egocentric left/right could be motivated by its applicability across environments. For this, I combine VR with the ‘experimental semiotics’ approach, where the game is solved via a novel visual communication channel. I show how the movement data in the 3D world can be correlated with invented signals to measure which FoR participants rely on. In contrast to the English data, I did not find an advantage for geocentric systems in the slope environment, and overwhelmingly egocentric systems emerged. I discuss how this could relate to task-specificity and native language background. More generally, I show how this new way of studying spatial language with interactive VR games can be used to test hypotheses about linguistic transmission and material culture that could help explain the origins of the egocentric FoR system, which is regarded a fairly recent cultural innovation.

Taken together, the thesis comprises several studies testing the relationship between linguistic and environmental variables. Additionally, VR is presented as a novel tool to study spatial language in controlled large-scale settings complementing more traditional fieldwork. More generally, I suggest that VR can be used to study the evolution of language in complex, multimodal settings without sacrificing experimental control.

Acknowledgements

Chronologically, I must first thank Monica Tamariz, whom I had met at the inaugural IACS all the way back in 2014, and who encouraged me to contact Kenny Smith with my ideas. Kenny suggested “to have a chat” and, somehow Simon Kirby and him turned out to be open to my somewhat-out-there idea of using interactive virtual reality in the lab – mind you, this was January 2015, when commercial VR was around the corner, but not just there *yet*. Simon was so enthusiastic that I only realised that this ‘chat’ was in fact an interview when I received a formal offer for the position two days later. Thinking about the application stage, I also wish to thank my former teachers in Aarhus, Kristian Tylén and Svend Østergaard, for encouraging me to turn this idea into a PhD. I am extremely sad that Svend did not live to see the VR experiments. He was a wonderful mentor, and I’ll never forget his support, and particularly his words “this project has to happen!”.

It took me a year to acquire funding, but in a way it was better to start in 2016, since VR technology had improved and the LEC had just evolved into the CLE and could support me in buying the lab equipment, for which I am extremely grateful. The CLE was what made this PhD both enjoyable and possible. I cannot imagine a better team of supervisors than Kenny, Simon and Jennifer Culbertson. My project certainly wasn’t the standard and it’s frankly amazing that doing a PhD in a linguistics department can consist of a year of learning C#, 3D-modelling and figuring out how on earth two VR headsets could possibly be linked up over a network to study human interaction. All the more, I want to thank them for their guidance (and patience!) as well as personal support that went far beyond what is normally expected from academic advisors. I know this wasn’t the easiest supervision, and I took nothing for granted, particularly the extra one-on-one sessions and support during the pandemic.

In a similar vein, my examiners Jürgen Bohnemeyer and Chris Cummins managed to make my examination both challenging and enjoyable by asking relevant, curious and insightful questions that led to very fruitful discussions and made me think more deeply about my research.

The fact that I conducted the ambitious experiments that I had initially planned, is a testament to the CLE as a top-notch research environment. In addition to my supervisor trio, I learnt a lot from the other PhDs, including those from my cohort, Andres Karjus, Ash Sato, Fausto Carcassi, Fiona Kirton, José Segovia Martin, and of course the other students, Annie Holtz, Cathleen O’Grady, Carmen Saldana, Henry Conklin-Coxe, Jasmeen Kanwal, Jon Carr, Marieke Woensdregt, Svenja Wagner, Tamar Johnson and Yasamin Motamedi, all of which made the PhD retreats a real treat. Special thanks to Andres for R-

troubleshooting sessions, and Fausto for his mathematical insight and sharing the pain not only as an officemate, but also as a flatmate.

Many thanks also to all the external collaborators involved in the projects that are listed in this thesis: Peeter Tinitis, Stefan Hartmann, Michael Spranger, Kristian Tylén, Riccardo Fusaroli and Marlene Staib. Research is a collaborative effort and I could not have produced any of the work without them. I would also like to thank the spatial linguistics community from whom I learnt a lot as somewhat of an outsider to the field. The theoretical backbone of the chapters on spatial referencing received a lot of calcium in discussions with Bill Palmer, Jürgen Bohnemeyer, Alice Gaby, Jonathan Schlossberg, Jonathon Lum, Steve Levinson and – of course – Martin Thiering, the original sinner, who introduced me to cognitive science and spatial language. The VR experiments reported here would also not have been possible without the help of Paola Peña, Elliot Huggett, Jasmine Salmon, Lucie Wolters and Hyowon Kwon, who assisted me with piloting and data collection. I would also like to thank the PPLS school at Edinburgh for supporting my work with a Principal’s Career Development scholarship and the school’s support staff, particularly Katie Keltie, Toni Noble, Cedric Macmartin, Steven McGauley, Susan Hermiston, Henry Barnett, Alisdair Tullo, and many others behind the scenes. In the CLE, I also very much enjoyed discussions with Chrissy Cuskley, Matt Spike and Marieke Schouwstra (and basically everyone else who worked or visited in the postdoc office between 2016 and 2020). I’d also like to thank Hannah Rohde for being a great course organiser and giving me so many opportunities to be involved with teaching and designing course materials over the last four years.

Outside of Edinburgh, I’d also like to thank my co-organisers of various academic events that often directly inspired thoughts in this thesis: Roberta Rocca, Mark Dingemanse, Thomas van Hoey (and again Michael Spranger, Stefan Hartmann and Peeter Tinitis). Thanks also to the vibrant evolutionary linguistics community that make conferences so enjoyable, including Niklas Johansson, Tessa Verhoef, Hannah Little, Alan Nielsen, James Winters, Thomas Müller, Gareth Roberts, Seán Roberts, Nic Fay, Bruno Galantucci, Vinicius Silva Macuch, Limor Raviv and many many others.

I am also indebted to Josh Cox who proofread many chapters of this thesis, partly under tight time-constraints. Any remaining errors are, of course, my own.

Lastly, I cannot thank my family enough for supporting me at all stages of my education over all these years, many of which I spent far away and definitely not visiting often enough (especially during the final year). Vielen Dank für alles! And finally, Hyowon. You are 6000 km away from me as I type this and yet no single word in this thesis would exist without your continuous support. 모든 것에 감사합니다.

Contents

1	Introduction	1
1.1	Preface to Paper 1	3
1.2	Paper 1: Language evolution research in the year 2020: A survey of new directions	4
	Abstract.....	4
	1. Introduction.....	4
	2. <i>Quo vadis</i> language evolution research? An overview of new directions.....	5
	2.1 Multimodality and iconicity.....	6
	2.2 An integrative approach to language dynamics	8
	2.3 Methodological innovations.....	10
	3. Conclusion	15
	Acknowledgments.....	15
1.3	Summary of Paper 1.....	15
1.4	Linguistic adaptation	16
1.5	An experimental evolutionary approach for investigating linguistic adaptation	21
1.6	Roadmap.....	23
2	Linguistic adaptation in the Laboratory.....	25
2.1	Preface to Paper 2.....	25
2.2	Paper 2: The emergence of systematicity: how environmental and communicative factors shape a novel communication system	26
	Abstract.....	26
	1. Introduction.....	26
	1.2 Outline of the study	29
	2. Experiment 1: Systematicity is motivated by structural properties of the environment.....	31
	3. Experiment 2: Systematicity emerges in displaced interactions and expanding meaning spaces	40
	4. General discussion	45
2.3	Summary of Paper 2.....	50
3	Spatial language: Linguistic adaptation at the intersection of language, culture and cognition.....	53
3.1	Introduction: Spatial language and cognition.....	53
	3.1.1 Frames of Reference.....	54
3.2	Origins and diversity of FoR.....	57

3.3	Preface to Paper 3.....	61
3.4	Paper 3: From the field into the lab: Causal approaches to the evolution of spatial language	
	62	
	Abstract	62
	1 Introduction.....	62
	2 Causality and limitations of observational data.....	64
	3 Addressing causality with models & experiments.....	68
	4 Conclusion.....	72
	Acknowledgements.....	74
3.5	Summary of Paper 3	74
4	Investigating the adaptivity of spatial language in virtual reality.....	75
4.1	Introduction	75
4.1.1	A large-scale virtual reality approach.....	75
4.2	Experiment 1: FoR adaptivity in English.....	80
4.2.1	Outline and prediction.....	80
4.2.2	Methods.....	81
4.2.3	Results	92
4.2.4	Discussion.....	103
4.3	Experiment 2: FoR flexibility when switching environments.....	105
4.3.1	Outline and prediction.....	105
4.3.2	Methods.....	105
4.3.3	Results	107
4.3.4	Discussion.....	109
5	Emergence of a novel spatial communication system in the lab.....	111
5.1	Introduction to experiment 3	111
5.2	Participants.....	112
5.3	Materials	112
5.4	Procedure.....	115
5.5	Results.....	115
5.5.1	Performance	115
5.5.2	Qualitative analysis of the communication medium.....	117
5.5.3	Frames of Reference inferred from movement data.....	120

5.6	Discussion.....	134
6	General discussion.....	137
6.1	Implications of the OrbHunt experiments.....	137
6.1.1	FoR and task-specificity.....	137
6.1.2	Potential follow-ups and future experiments.....	140
7	Concluding remarks.....	143
7.1	Recapitulation.....	143
7.2	Final thoughts.....	144
	Appendices.....	146
	Appendix A: Data and code experiments 1-3.....	146
	Appendix B: OrbHunt: Setup and technical specifications.....	147
	Appendix C: Instructions for experiments 1-3.....	149
	Appendix D: Post-test questionnaire for experiment 3.....	152
	References.....	153

1 Introduction

Every subject spins out, like the spider's threads, its relations to certain qualities of things and weaves them into a solid web, which carries its existence.

— Jakob Johann von Uexküll (1934: 327)

Why are languages structured in the way they are and why do they vary in the way they do? There is striking variation in semantic categories (e.g., colour, space), morphosyntax (e.g., word order, gender and case marking), spoken or signed phonology (e.g., gesture or sound inventories, phonotactics) and even pragmatics, how speech is integrated with non-verbal signals, how social hierarchies are expressed grammatically, and so on (For an introduction to linguistic diversity see Hammarström 2016; Haspelmath et al. 2005). Where do these differences come from? Are they the simple outcome of random drift and arbitrary cultural choices? Or could they be systematically motivated by external factors?

In this thesis, I investigate whether specific differences in linguistic structure that we find across different languages might relate to the wider external environment in which these languages are embedded and have historically evolved in. In other words, the subject of this thesis is 'linguistic adaptation', the idea that languages, just like biological organisms, can adapt to their local environment or ecology (Lupyan & Dale 2016). While in biology, 'environment' usually refers to a particular ecological niche or habitat, in the case of languages, these can encompass a variety of contexts that language is embedded in. On the one hand, there is the natural ecology of language, that is, its primary use in multimodal and direct face-to-face interaction, where speech and gesture are used for information transfer to coordinate with an interlocutor and achieve common communicative goals (Clark 1996; Fusaroli & Tylén 2012; Holler & Levinson 2019). As they are used for communication and transmitted to new generations, languages are thus constantly shaped to be as simple, learnable, and expressive as possible, to enable us to communicate in efficient ways about the world (e.g., Kemp & Regier 2012; Kirby et al. 2015). On the other hand, the world itself can in turn have an impact on language. Specific referential contexts might alter the relevance of semantic features in an utterance and thus its degree of specificity (e.g., Koolen, Goudbeek & Kraemer 2013), and specific environments might favour different linguistic solutions even on a grammatical, lexical or semantic level. This can go as far as including physical aspects of the environment such as climate (e.g., Everett, Blasi & Roberts 2015) or topography (e.g., Palmer et al. 2017), social aspects, such as education or contact with other languages (e.g., Bohnemeyer et al. 2015), and even technologies like writing that can affect utterance production (e.g., in prosaic writing vs written chats, see Lupyan & Dale 2016: 8).

How can we study to what extent and through which mechanisms the physical or sociocultural environment affects language? I suggest that for addressing such fundamental questions about our unique communication system, we have to look at it through an evolutionary lens (see, e.g., Hurford, Studdert-Kennedy & Knight 1998). On an evolutionary view, languages can be considered the outcome of the forces of self-organisation and selection (Steels 2011) that constantly operate as languages are used, learnt and transmitted within and between human populations. Horizontal interaction with other speakers or signers and vertical transmission to new generations of learners exert complementary pressures that affect languages in different ways (e.g., Smith, Tamariz & Kirby 2013; Carr et al. 2020; Kirby et al. 2015; Christiansen & Chater 2008; Christiansen & Chater 2016). I argue that besides internal cognitive and linguistic biases, language never evolves in isolation. It is embedded in sociocultural settings, tied to technologies and used to communicate about the physical world in a way that is meaningful and relevant to its users. In this sense, language forms a cybernetic feedback loop with its environment, or what the bio-semiotician Jakob von Uexküll (1992) called 'Umwelt'. This means that like biological organisms that adapt to their ecological niche, it is not merely variation and neutral cultural drift that makes languages drift apart and leads to linguistic diversity, but also selection through functional adaptation of linguistic features. This view of language evolving in a 'Darwinian' sense is not new and has been discussed at least since the days of Schleicher (1863), Müller (1870), and Darwin (1871) himself. However, in modern language evolution research, it has only recently become more popular again, due to the availability of larger amounts of cross-cultural data having led to a surge in studies exploring the relationship between linguistic and non-linguistic variables to uncover statistically robust patterns (Roberts & Winters 2012; Ladd, Roberts & Dediu 2015), such as, e.g., the famous case of the inverse relationship between morphological complexity and population size (Lupyan & Dale 2010).

The aim of this thesis is to complement this line of research with what one could call the 'language-evolution-in-the-lab approach'. In this approach, controlled laboratory experiments with human subjects are used to simulate the emergence and evolution of linguistic building blocks or conventions in a controlled laboratory setting, which allows formulating and testing hypotheses about the conditions that give rise to these structures and how they are shaped by learning, transmission and use (see, e.g., Galantucci, Garrod & Roberts 2012; Kirby, Griffiths & Smith 2014; Tamariz 2017).

In this first chapter, I will first present a paper that provides a very general overview about the interdisciplinary field of language evolution research (Paper 1, Nölle, Hartmann & Tinitis 2020). In section [1.4](#), I will provide more details and examples for linguistic adaptation. Based on the literature reviewed, I will argue that to fully understand the relationships between linguistic and non-linguistic variables, cross-

sectional observations are not enough, and instead causal hypotheses should be tested in the laboratory. Section [1.5](#) will briefly review existing experiments that have been successfully used to study the cultural evolution of language in the lab and uncovered some of the pressures that lead to the emergence of fundamental linguistic design features like compositionality or combinatoriality. I will also discuss recent studies that have considered language-external variables, like the impact of different social dynamics or referential contexts. In the same vein, the original research presented in this thesis will apply the same experimental approach to study the role of linguistic adaptation in cultural language evolution. Finally, the present chapter will conclude with a brief roadmap that outlines these research projects, which are detailed in chapters 2, 4 and 5.

1.1 Preface to Paper 1

Paper 1 (Nölle, Hartmann & Tinitis 2020) constitutes the introduction to a special issue on “New directions in language evolution research” in *Language Dynamics and Change*, which I co-edited with Stefan Hartmann (University of Düsseldorf) and Peeter Tinitis (University of Tartu). The contributions comprise a diverse selection of interdisciplinary research into the origins and evolution of linguistic structure including perspectives from biology, archaeology, sociolinguistics, historical linguistics, sign language research, and cognitive ethnography and were published in two separate volumes of the journal in 2020. The perspective paper was written by all three of us in collaboration, with me serving as lead author. Additionally, the paper provides an overview about recent trends, methodological innovations, and the direction the field is headed in. In the context of this thesis, it shall serve as a brief introduction to the diverse field of language evolution research.

1.2 Paper 1:

Language evolution research in the year 2020: A survey of new directions

Jonas Nölle, Stefan Hartmann, Peeter Tinitis

Abstract

This introductory paper reviews recent advances in language evolution research and summarizes the contributions of the special issue “New Directions in Language Evolution Research” in the broader context of these developments. Specifically, we discuss the increasing role of multimodality and iconicity, the more integrative view of language dynamics that has arguably broadened the scope of language evolution research, and recent methodological innovations that allow for a more fine-grained study of e.g. typological distributions or behavioral patterns that can give clues to some of the key questions discussed in the field.

1. Introduction

Research on language evolution is undoubtedly among the fastest-growing topics in linguistics. This is not a coincidence: While scholars have always been interested in the origin and evolution of language, it is only now that many questions can be addressed empirically. Researchers can now draw on a wealth of data and a multitude of methodological approaches that have been developed in the different disciplines that try to find answers to what has been called “the hardest problem in science” (Christiansen & Kirby 2003). Importantly, any theory of how language may have emerged requires a solid understanding of how language and other communication systems work. As such, the questions in language evolution research are not just manifold but also interface with key open questions in historical and theoretical linguistics: Where does human language come from and what makes it unique compared to animal communication systems? How do cognition, communication and transmission shape grammar? Which factors can explain linguistic diversity? How and why do languages change? To what extent is the structure of language(s) shaped by language-internal and external environmental factors?

The aim of this introductory article is to give a brief overview of current new directions in language evolution research that are taken to address these questions. As Dediu & de Boer (2016) have noted, the field of language evolution research is currently coming of age, and has developed a rich toolkit of methods both for comparative research, which investigates the commonalities and differences between human language and animal communication systems, and for studying the cumulative cultural evolution of sign systems in experimental settings, including both computational and behavioral approaches (see, e.g.,

Tallerman & Gibson 2012; Fitch 2017) In addition, large-scale typological studies have gained importance in recent research on language evolution (e.g., Evans 2013; Blasi et al. 2016; Majid et al. 2018; Jackson et al. 2019). Here, we survey important recent developments in the field, and discuss how the papers collected in this special issue of *Language Dynamics and Change* can be seen as representative of these new directions. Note that the special issue is split into two volumes, which is why some of the papers summarized here will appear in the next issue of LDC.

2. *Quo vadis* language evolution research? An overview of new directions

The view that there has been a “gap” in language evolution research between the infamous ban of discussions relating to the emergence of language by the *Société de Linguistique de Paris* in 1861 and an alleged resurgence of the field in 1990s has long been identified as a myth (see, e.g., Fitch 2010: 16). However, it is true that the field has seen a surge since the late 20th century, and it has evolved considerably over the past two decades. In this section, we briefly outline a number of major trends that are also taken up in the contributions to the special issue: the increasing importance of multimodality and iconicity in accounts of the evolution of language; taking into consideration contextual and environmental factors more thoroughly; and the growth of interdisciplinary approaches in both theory and method.

Regarding the relationship between theory and method, the question of causality has been discussed intensively. Various interacting factors have led to a shift in perspective here: First, the notorious problem of spurious correlations has been pointed out in various papers (e.g., Roberts & Winters 2012; Roberts & Winters 2013). Second, the “replication crisis” (e.g., Open Science Collaboration 2015) has affected virtually all domains of empirically-based scientific inquiry, including linguistics, and third, language evolution research often relies on very indirect evidence. This makes it attractive not only for theoretical “just-so stories” but also for both far-reaching and far-fetched empirical hypotheses, which increases the problem of spurious correlations and makes the sanity checks that have been called for in the wake of the replication crisis all the more necessary.

The key question, then, becomes: Which hypothesized causal relationships are robust and what kind of evidence can support those causal links? Many of the variables discussed in language evolution research are causally linked and interact with each other in different ways. Therefore, in order to avoid the problem of spurious correlations, a new approach, which S. Roberts (2018) terms the ‘maximum robustness approach’, is gaining momentum. Instead of focusing on simple relationships, e.g., environmental variable X correlating with linguistic trait Y, the idea is to systematically construct more complex and coherent causal

graphs (Pearl 2009) that are based on all the available evidence as well as hypothesized links between multiple interrelated variables. CHIELD, a novel database for hypotheses in language evolution, allows for exploring such graphs and identifying complementary or conflicting relationships as well as empirical gaps that lack evaluation, which can help design empirical research addressing those issues and uncovering actual causal mechanisms (Roberts 2018; Roberts et al. 2020).

2.1 Multimodality and iconicity

The modality where language originated has been hotly debated: According to the “gesture-first” view, “language evolved initially from manual gestures with vocal elements gradually added” (Corballis 2012: 383; see also, e.g. Corballis 2003; Tomasello 2008; Arbib 2012 among others; McNeill 2012 for an overview and criticism). The “speech-first” view defended by e.g. Dunbar (1996) and MacNeilage (2008), by contrast, argues that language must always have existed in the vocal-auditory modality given its present-day dominance (Zlatev et al. 2017: 453). These different theories, which go back even further than the works cited here (see, e.g., Fitch 2010 for a historical overview), are now increasingly replaced by more nuanced and fine-grained views that argue for a complex interplay between auditory and visual channels in the early multimodal beginnings (e.g., Kendon 2017; Perlman et al. 2017). For example, regarding the visual-spatial modality, the idea of “pantomime-first” has been put forward as “a theoretical proposal that is arguably distinct from both gesture-first and multimodal-first” (Zlatev et al. 2017: 473; also see Żywicznyński, Waciewicz & Sibierska 2018).

In line with the strongly empirical outlook of modern language evolution research, a number of studies have set out to test the predictions of those different theories empirically (Zlatev et al. 2017; Macuch Silva et al. 2020). Across modalities, the key role of iconicity has received increased attention (Perniss, Thompson & Vigliocco 2010; Perniss & Vigliocco 2014; Dingemanse et al. 2015; Pleyer et al. 2017) with the similarities and differences between modalities in the affordances for iconicity being highlighted (Perlman et al. 2018) and methodological innovations developed to assist systematic study (Motamedi et al. 2019a).

Importantly, these theories go in tandem with a more general development in the language sciences noted by e.g. Waciewicz & Żywicznyński (2017), namely that face-to-face interaction represents “the core ecological niche for language” (Torreira, Bögels & Levinson 2015: 1). Language is therefore more and more regarded as a multimodal phenomenon. In particular, the role of gesture has received increasing interest (Müller et al. 2013). The affordances of the gestural modality for grounding communication systems have been investigated quite extensively. On the one hand, experimental studies have shown how

participants develop new gestural communication systems from scratch (Galantucci, Garrod & Roberts 2012). On the other hand, the emergence of new sign languages such as Nicaraguan Sign Language (NSL) and Al-Sayyid Bedouin Sign Language has supplemented theories of language emergence with natural data.

Carla Hudson Kam and **Oksana Tkachman** add to this line of research by investigating the role of interpretability. In particular, they focus on the use of spatial modulations to encode concrete spatial relations on the one hand and more abstract grammatical relations on the other. Starting from the finding that the way entities are placed in signed space seems to facilitate the way grammatical relations, but not actual spatial ones, are understood in NSL, they conduct a series of experiments with hearing non-signers – adults and children – testing the hypothesis that the locative use of space is less directly interpretable as it requires more interpersonal coordination and thus takes more time to conventionalize. In a first experiment, participants were presented with a short narration mentioning two characters, with a pronoun referring back to one of them. Depending on the condition, the narrator used no gesture, the same gesture for the character and the pronoun referring to this character, or different gestures for the character and the pronoun. Hudson Kam and Tkachman's results show that both children and adults interpret pronouns differently based on co-speech gestures. In a second experiment, participants also watched short narratives involving two characters, with the narrator producing co-referential localizing gestures indicating the spatial location of the two characters (left vs. right). In a picture selection task, participants were then asked to indicate which picture best described the story they had heard. They were shown two versions of the same scenario that differed in the relative spatial location of the characters. While the adults showed some limited consistency in their interpretation, the children did not. They conclude that adults and older children, even if not familiar with a signed language, take in spatial information conveyed via co-speech gesture while younger children do not, possibly because the interpretation of spatial information is quite demanding in that it requires inferences about spatial configurations as well as about perspectivation.

Whereas Hudson Kam & Tkachman are concerned with integral features of sign languages that do not arise in spoken language, the study by **Katie Mudd, Connie de Vos & Bart de Boer** uses data from an emerging sign language for modelling language persistence. Building on a mathematical model of sign language persistence (Aoki & Feldman 1991), they inform their agent-based model with real-world data. In particular, they investigate the impact of extralinguistic factors such as the number of deaf individuals in the population, the number of hearing carriers of the deaf allele, population size, and mating patterns. In their model, the sign language does not persist in a small population, while the proportion of deafness in the population only leads to non-persistence if it reaches an extremely low threshold. However,

they show that for both results, there are counterexamples in existing emerging sign languages, and they discuss additional factors that may be necessary to explain the persistence of sign languages.

2.2 An integrative approach to language dynamics

Another important trend in language evolution research is certainly that its scope has extended quite significantly. This can partly be seen as a by-product of a different but related development, namely the increasingly widespread view of language as a complex adaptive system, i.e. a system whose global properties arise from multiple independent interactions on the local level (e.g., Steels 2000; Beckner et al. 2009; Kirby 2011), which entails a more unified approach to the ontogenetic, phylogenetic, and “glossogenetic” (Hurford 1990) dimensions of language dynamics. This is why the term “language evolution” is sometimes used without terminological consistency, as Haspelmath (2020) notes: On the one hand, it can refer to the biological evolution of the biological capacity for language (called “linguisticity” by Haspelmath), on the other hand, it can refer to the cultural evolution or diachronic change of particular language systems (also see Mendívil-Giró 2019; Hartmann 2020). While Haspelmath (2020) argues that “[t]he comparison of languages does not lead to immediate insights about human linguisticity”, various scholars have defended the view that investigating present-day distributional patterns can yield valuable clues to the emergence of language. For instance, Heine & Kuteva (2007; 2012) argue that “typological generalizations on language structures” (Heine & Kuteva 2007: 55) can be used to reconstruct the “genesis of grammar” with the help of grammaticalization theory. In a similar vein, Bybee (2010: 202) emphasizes the relevance of investigating historical language change for understanding language evolution: “there is every reason to suppose that the very first grammatical constructions emerged in the same way as those observed in more recent history”. However, note that both Bybee and Heine & Kuteva hedge their hypothesis carefully, emphasizing that their accounts can only give clues to the origins of grammatical units, not language per se. However, as e.g. **Antonio Benítez-Burraco** points out in his paper, investigating both typological and grammaticalization patterns can point to environmental and social driving forces in cultural evolution, which in turn can indeed be relevant for understanding the origins of the human capacity for language.

While Haspelmath’s (2020) distinction between “evolution of linguisticity” and “evolution of languages” reminds us that both biological and cultural factors are at play in language development, it is also important to keep in mind that we cannot clearly tease apart the emergence of the language capacity from the cultural evolution of language(s). This point is made in a particularly forceful way by coevolutionary models, which assume that the evolution of language, culture, and biological features interact (see, e.g., Evans 2013; Kirby 2017) In the present special issue, Benítez-Burraco makes a strong case for this

hypothesis. Arguing against the view that “the evolution of language” on the one hand and “the evolution of languages” on the other can and should be teased apart, he points to recent evidence according to which (i) human cognition might not be as uniform as is usually assumed and might have changed over our recent history, and (ii) the structure of languages is partly shaped by social, cultural, and environmental factors. He therefore argues for an eco-evo-devo account of language evolution, emphasizing the role of human self-domestication, i.e. the emergence of traits typically associated with domestication in biological evolution, but without external influence.

Studies investigating historical language change have shown that methods from evolutionary biology can be fruitfully applied to the reconstruction of language history (see, e.g., Gray, Drummond & Greenhill 2009; Syrjänen et al. 2016; Bouckaert, Bowern & Atkinson 2018). Additionally, theoretical approaches like Croft’s (2000) evolutionary approach to language change or Blevins’ (2004) and Ritt’s (2004) evolutionary approaches to phonology have shown the potential of mutual cross-fertilization between linguistics and evolutionary theory. Language itself has come to be viewed more holistically. On the one hand, as pointed out in Section 2.1, language is increasingly seen as a fundamentally multimodal phenomenon. On the other hand, despite its undisputed special status, language is no longer seen as the “odd one out” among our cognitive and cultural traits. Instead, an integrative view sees language as bearing important commonalities with other phenomena, which is why phenomena as different as the construction of paper airplanes (Caldwell & Millen 2008), the evolution of a collaborative art project (Müller & Winters 2018), diachronic changes in ancient coin designs (Pavlek, Winters & Morin 2019), and the evolution of language(s) can be investigated in a shared theoretical framework. The 2016 founding of the Cultural Evolution Society bears witness to the appeal of such an integrative approach to language, cognition, and culture.

These ideas are of course not entirely new. Among others, Humboldt (1836) has already argued for a close connection between language, cognition, and culture, even though many of his ideas are by now outdated, to say the least (e.g., the assumption that the complexity of a language reflects a nation’s degree of civilization). What is (relatively) new, however, are many of the methods that can now be used to study these relationships in more detail. This arguably also entails a more thoroughly interdisciplinary outlook: While language evolution research has been a highly multi-disciplinary field from the outset, it could be argued that this interdisciplinarity has often consisted in combining insights from studies that themselves largely remained within disciplinary boundaries. Many more recent studies, however, cross these boundaries by taking a strongly interdisciplinary approach already in the design of their studies. For example, Dediu & Levinson (2013; 2018) argue that Neanderthals had vocal language on the basis of archaeological

as well as genetic evidence, and Tylén et al. (2020) combine the analysis of archaeological findings with behavioral experiments to investigate the evolution of symbolic communication.

The interpretation of archaeological findings for studying language evolution entails some obvious challenges: For instance, it is a matter of debate what kinds of archaeological records constitute evidence for symbolic or even linguistic behavior (see, e.g., d'Errico & Vanhaeren 2009; Botha 2012). This is why **Cory M. Stade** proposes to use Theory of Mind (ToM) as a proxy for language ability, rather than relying on symbolism. She argues that ToM might be better identifiable in the archaeological record as certain social behaviors have been shown experimentally to require theory of mind. She mentions a number of advantages, e.g. the strong correlation that has been found between ToM and language, which has led to the widely-shared assumption that both have co-evolved (see Woensdregt, Cummins & Smith *forthc.* for a computational model). Stade's study is particularly relevant in light of recent discussions about ToM in general (see, e.g., François & Rossetti 2020; Sulik & Lupyan 2020) and may give rise to methodological discussions about how ToM can be defined and operationalized more thoroughly.

Representing the explorations of interdisciplinary connections, **Gareth Roberts and Betsy Sneller** make a case for increased collaboration between sociolinguists and language evolution researchers by comparing the five foundational problems in language change (Weinreich, Uriel, Labov & Herzog 1968) and the four foundational questions in ethology (Tinbergen 1963). The paper discusses resemblances between sociolinguistic and cultural evolutionary approaches that both deal with population-level problems like the spread and propagation of features as well as opportunities for integrating them in a mutually informative way. Roberts and Sneller further show how experimental paradigms that are usually used to study language evolution can be used to test sociolinguistic theories of language change, and how the study of natural language change can in turn inform experiments in cultural language evolution. An interesting point that they note is that the questions driving language evolution research today were already embraced by Weinreich, Labov & Herzog in their formulation of a general science of language change.

2.3 Methodological innovations

In his recent detailed review, Fitch (2017) already lays out the wide varieties of methods and interdisciplinary approaches that are now available to researchers from various fields to study the components of the human language faculty, drawing on comparative biology, neuroscience, paleontology and genetics. Similarly, there has been an explosion in experimental studies addressing the cultural evolution of linguistic structure in the lab, ever since pioneering studies showed that both the grounding of communication systems from scratch in horizontal interaction (Galantucci 2005), and the vertical cumulative cultural

evolution of language over generations (Kirby, Cornish & Smith 2008) can be simulated with human participants in controlled settings. These approaches are now referred to as ‘experimental semiotics’ (Galantucci, Garrod & Roberts 2012) and ‘iterated learning’ (Kirby, Griffiths & Smith 2014). These approaches initially served as empirical validations of early computational models that suggested linguistic structures could emerge through interaction and transmission (e.g., Steels 1995; Kirby & Hurford 2002). However, since then, studies using artificial communication systems have matured into a range of paradigms that have been used to study the emergence and evolution of linguistic structure and the pressures shaping it (Tamariz 2017). General observations include different pressures for simplicity/learnability and expressivity/informativeness that can result from horizontal interaction or vertical transmission (Kirby et al. 2015; Motamedi et al. 2019; Carr et al. 2020). Other recent studies have scrutinized in what way communication systems are sensitive to properties of the signaling space (Little, Eryılmaz & de Boer 2017; G. Roberts, Lewandowski & Galantucci 2015), properties of the meaning space (Carr et al. 2017), the referential context (Winters, Kirby & Smith 2015; Winters, Kirby & Smith 2018; Tinitis, Nölle & Hartmann 2017), factors related to the mode of communication (Nölle et al. 2018b), or information-theoretic principles (Kanwal et al. 2017; Stevens & Roberts 2019). A general trend seems to be that experiments are not only used to study highly abstract phenomena any more, such as the emergence of basic design features like compositionality, but also cognitive biases and cultural evolutionary mechanisms that affect the distribution of concrete linguistic features found among natural languages (Roberts 2017). This includes, for instance, how languages order semantic constituents (e.g., Goldin-Meadow et al. 2008; Schouwstra & de Swart 2014; Özçalışkan, Lucero & Goldin-Meadow 2016; Meir et al. 2017) or modifiers within the noun phrase (Culbertson & Adger 2014; Martin et al. 2019) as well as how the social environment, e.g., group size (Raviv, Meyer & Lev-Ari 2019) or social biases (Sneller & Roberts 2018; Roberts & Fedzechkina 2018), affect the evolution of linguistic systems. In this way, artificial language learning, communication and transmission studies have become powerful tools to study the mechanisms underlying linguistic universals and diversity more generally. This point is echoed by G. **Roberts and Sneller**, who argue that artificial language experiments could have a significant impact on sociolinguistics and studies of language change.

Besides cross-fertilization with other linguistic subfields, multidisciplinary approaches to language evolution are now commonplace. This is in part due to the realization that language does not evolve in a void but is rather “the product of a multiscale communicative niche construction process at the intersection of biology, environment, and culture” as Coupé et al. (2019: 1) formulate it. There is now increasing evidence that a large number of features that vary widely across languages, such as specific speech sounds

(e.g., Everett, Blasi & Roberts 2015; Dediu, Janssen & Moisik 2017) or lexical or semantic categories (e.g., Brown & Lindsey 2004; Regier, Carstensen & Kemp 2016) might not just reflect arbitrary cultural ‘choices’, but are rather motivated by and sensitive to the specific local physical, sociocultural or technological environment (see Lupyan & Dale 2016 for an overview). While this kind of ‘linguistic adaptation’ was initially suggested due to observed correlations, such as the inverse relationship between morphological complexity and population size (Lupyan & Dale 2010), such findings are now corroborated by phylogenetic evidence suggesting that the historical diversification of language families is better explained by environmental factors such as climate rather than neutral drift (Bentz et al. 2018; Hua et al. 2019). An ambitious paper by Blasi et al. (2019) illustrates just how rigorous and transdisciplinary more recent approaches to linguistic adaptation have become: Combined results from biomechanical modeling of bite configurations with phylogenetic analyses of language change and ethnographic data suggests that changes in diet during the neolithic revolution have affected the distribution of labiodental sounds such as [f] and [v]. These phonemes are significantly easier to produce with an overbite that resulted from the transition into agriculture leading to a softer diet.

Such endeavors are to some extent enabled by a recent surge in large-scale databases enabling what Bickel (2015) calls ‘distributional typology’, a data-driven comparative approach that addresses the *what*, *where* and *why* questions regarding language structure with statistical techniques. With Glottolog (Hammarström, Forkel & Haspelmath 2019), linguists have access to an accurate classification for approximately 6500 – 8000 linguistic varieties including information on (sub-)families and dialects (see Hammarström 2016 for an introduction and overview) and there are now many specialized databases for linguistic features (e.g., WALS, Dryer & Haspelmath 2013), phoneme inventories (PHOIBLE, Moran, McCloy & Wright 2014), basic vocabulary word lists (ASJP 2020) as well as databases that allow linking linguistic and cultural or environmental data (D-PLACE, Kirby et al. 2016). These and other data sources (see <http://languagegoldmine.com/> for a comprehensive list) are now increasingly used to investigate general patterns found across languages, such as systematic sound-symbolic meaning associations in the lexicon (Blasi et al. 2016) or universals and variation in colexification of emotion words across thousands of languages (Jackson et al. 2019). Similarly, an ever-growing number of corpora enables detailed synchronic and diachronic analyses of single languages and allows characterizing individual- and population-level dynamics that are involved in historical language change (e.g., Petré & Van de Velde 2018). Recent work using English corpora has, for instance, addressed to what extent sensory vocabulary reflects communicative efficiency across semantic domains (Winter, Perlman & Majid 2018).

The paper by **Andres Karjus, Richard A. Blythe, Simon Kirby, and Kenny Smith** represents another example of this line of research. They discuss how large-scale corpora can help understand language dynamics and change. Previous work in the field of “culturomics” (e.g., Michel et al. 2011) has been heavily criticized because “these approaches often present sweeping accounts of the historical development of words and concepts with little or no attention paid to important linguistic clues and principles of good science.” (Brezina 2018: 224) Unlike those previous works, the “topical-cultural advection model” that Karjus et al. propose aims at teasing apart different explanatory factors in order to separate random from systematic effects. Two case studies show that the model is able to capture topical effects in frequency changes of linguistic elements over time.

With the availability of rich data sources it is tempting to simply search for cross-sectional patterns. For instance, economists have recently begun linking linguistic features such as future-tense marking or pronoun-drop to economic behavior (e.g., Chen 2013; Kim, Kim & Zhou 2017; Feldmann 2019). However, one argument for the ‘maximum-robustness’ approach discussed in section 2 above is that such large-scale correlational studies are not well-suited to make causal claims, since relationships between linguistic and language-external variables are often complex and confounded (e.g., due to historical or geographic relatedness of cultural traits). It has therefore been argued that experiments or case studies are a better way for verifying such relationships (Roberts, Winters & Chen 2015; Roberts 2018b).

This is exemplified by the relationship between spatial language and the environment: For a long time, field linguists have speculated that the way in which speech-communities reason and talk about spatial relations could be dependent on the local topographic environment (e.g., Wassmann & Dasen 1998). However, large-scale surveys did not find simple relationships between spatial linguistic strategies and environmental factors such as climate or subsistence style (Majid et al. 2004). Only more recently, detailed individual case studies and comparisons between environments and speech-communities have found that social and environmental variables such as second language, subsistence style or local topography do indeed explain some of the variance in spatial linguistic strategies (e.g., Bohnemeyer et al. 2015; Palmer et al. 2017; Shapero 2017). However, it is hard to disentangle the exact causal relationships between such variables. **Magnus Phrao Hansen and Carolyn O’Meara** therefore argue for an integrated approach to the adaptation of spatial language, through a combination of ethnolinguistic fieldwork and historical studies of language change. They provide a detailed framework for selecting linguistic varieties with regard to their grammatical differences, ambient climate and landscape features, and available historical data that can be used to reconstruct whether an adaptation to environmental factors has taken place in the development of these varieties. Observations from real linguistic data could then be tested with

computational models (see, e.g., Spranger 2016) or human subject experiments testing the impact of the environment on linguistic conventions (e.g., Nölle, Fusaroli, et al. 2020).

Russell Richie, Matthew L. Hall, Pyeong Whan Cho, and Marie Coppola provide another example that shows why it is important to address hypothesized causal relations with a mixed-methods approach. Building on previous work from experimental semiotics as well as natural observation and computational modeling, they investigate the role of social network dynamics in language emergence. Using a dyadic gestural communication task, they investigate whether conventionalization differs between two conditions, one with a richly-connected and one with a sparsely-connected communicative network. In the richly-connected communicative network, each member of a group of four participants gets to interact with every other member, while in the sparse condition, one member is involved in all dyads, thus serving as a “hub”. They show that conventionalization is greater and grows more quickly in the richly-connected network when measured per unit time, i.e. per round in the experiment; however, the effect disappears when measured per interaction (the richly-connected network condition contains more interactions than the sparsely-connected network condition). A possible explanation is that there are different mechanisms of conventionalization at work: In the sparse network, the person that serves as a “hub” can set a standard that other participants tend to converge on; while in the rich network, participants have to negotiate whose utterances they will converge on, which is why participants in richly-connected networks do not conventionalize referring expressions faster, per interaction, than those in sparsely-connected networks. In contrast, in a rich network with more interactions, conventionalization happens faster than in the sparse network.

Another promising avenue is the use of novel technology. For a long time, experimenters had to navigate a trade-off between experimental control and ecological validity. However, with the availability of new methodologies like immersive virtual reality (VR), it is now becoming clear that it is possible to design tasks that are both tightly controlled and naturalistic (Peeters 2019). VR holds a lot of potential for investigating cases like spatial language, e.g., by comparing spatial linguistic strategies across simulated large-scale environments that resemble those that have been speculated to affect natural languages (see Nölle, Kirby, et al. 2020). As discussed in section 2, it is this interdisciplinary approach that can help address causality and model the actual mechanism underlying cross-linguistic variation more rigorously and incrementally. Lastly, artificial intelligence researchers have recently begun modeling the emergence of compositional structure using deep reinforcement learning algorithms with a similar approach to early evolutionary linguistic models (Kirby & Hurford 2002). In such models, compositionality emerges in multi-agent learning environments with different signal and meaning spaces to enable coordination (Lazaridou,

Peysakhovich & Baroni 2016; Mordatch & Abbeel 2017; Choi, Lazaridou & de Freitas 2018). While this trend in AI has begun independently, recent work is beginning to converge with findings in language evolution research, for instance by introducing generational transmission as a cultural evolutionary mechanism (e.g., Cogswell et al. 2019; Ren et al. 2019).

3. Conclusion

The papers in this special issue show that the field of language evolution research has evolved substantially in recent years. Various new approaches, both theoretical and methodological, have been developed that shed new light on the key questions of language evolution research. But the new directions we have sketched here have not only influenced the way we address these questions. Instead, we have also seen a shift in focus regarding the research questions that are asked: Traditionally, the question of *when, how, and why* the capacity for linguistic communication evolved has been at the centre of language evolution research. As we have seen, these questions are still important: For instance, archaeological and genetic evidence keeps providing new insights on the antiquity of language. More and more, however, the focus has shifted to uncovering the co-evolutionary dynamics of language, cognition, and culture. This has also entailed a reconceptualization of the scope of language evolution research, which now involves the study of language dynamics and change more generally from a cultural evolutionary perspective. The contributions to this special issue reflect this tendency. They also show the strongly interdisciplinary outlook of modern language evolution research in discussing how insights from previously neglected fields can contribute to our understanding of how language evolved and continues to evolve.

Acknowledgments

The papers collected in this special issue partly go back to presentations at the workshop “New Directions in Language Evolution Research” at the 51st Annual Meeting of the Societas Linguistica Europaea in Tallinn. We are grateful to all participants for stimulating presentations and discussions. Also, we would like to thank the LDC editors Simon Greenhill and Jeff Good for their guidance throughout the editorial process as well as the anonymous reviewers who have contributed substantially to the quality of this collection.

1.3 Summary of Paper 1

Paper 1 gave an overview of current trends and future directions in language evolution research. The key point in the context of this thesis is that over the past decades, the field has matured into a serious inter-

and transdisciplinary project that attempts to explain the past origins as well as ongoing evolution of human language from a wide variety of fields that range from archaeology to zoology and methods that include comparative animal studies, computational models and human subject experiments, among many others. In contrast to earlier, more theoretical accounts (e.g., Pinker & Bloom 1990; Hauser, Chomsky & Fitch 2002), language evolution research has become much more focused on empirical evidence and the testing of falsifiable hypotheses to address questions about the fundamental nature of human symbolic communication. Importantly, there has been increasing evidence that the cultural evolution of language, which interacts to some extent with the biological evolution of the human linguistic capacity, can be studied with regards to general principles such as cognitive biases, pressures from learning and use and language-external factors (see, e.g., Steels 2011; Kirby 2017 for theoretical accounts; and Blasi et al. 2019 for a concrete example). It is therefore becoming increasingly accepted that the structure of human languages is not just the simple outcome of a genetic specialization and core computational properties (cf. Berwick & Chomsky 2017), but rather continuously shaped as it is transmitted from mind to mind and used for communication in face-to-face interaction (Christiansen & Chater 2008; Steels 2011; Kirby 2017; Martins & Boeckx 2019). Section 2.3 of the paper already provided some examples that show that on top of this, the social or cultural environment in which communication and transmission take place, can likewise have an impact and have therefore received more attention in recent years. Below, I will review this evidence in some more detail.

1.4 Linguistic adaptation

Why don't all humans speak the same language? Languages can essentially be viewed as sets of conventions that cultural groups have consciously and unconsciously agreed on in order to solve recurrent communication and coordination problems that pertain to an endless range of communicative activities, like buying a coffee, doing taxes, writing a PhD thesis, and so on (see, e.g., Clark 1996; Lewis 1969; Dor, Knight & Lewis 2014; Tomasello 2014). As comes with most cultural phenomena, there are of course various ways of going about things. It should thus not be surprising that many languages have found different solutions to expressing thoughts and referring to the world that manifest in all structural aspects, including phonology, morphosyntax, semantics, pragmatics as well as differences tied to modality (signed, spoken, written), social background, or pathological conditions, such as language impairments (see, e.g., Ammon et al. 2004; Evans & Levinson 2009; Haspelmath 2010; Kirby et al. 2016; Hammarström 2016; Chondrogianni & Kwon 2019). While our uniquely symbolic communication system has increased our ability to transmit and accumulate cultural knowledge (Tomasello 2008; Richerson & Boyd 2008),

linguistic features themselves can be regarded as cultural ‘replicators’ that have even been found to approximate the phylogenetic tree of our species (Pagel 2009).

In linguistics, the vastly influential generative grammar program (e.g., Chomsky 1965; see Haegeman 1997 for an introduction) has traditionally put emphasis on trying to uncover universal rules and principles that govern all languages, with some scholars, like Chomsky and Pinker, famously arguing that from an outside perspective, all humans essentially speak the same language (see Pinker 1994: 232). However, over the last few decades, an enormous amount of comparative cross-linguistic research has challenged this assumption. Based on diversity across virtually all subdomains of language, Evans and Levinson (2009) argue that languages are fundamentally variable at all levels of organisation, including sound, meaning and syntax, as a result of undergoing cultural evolution. This is in line with a general trend in the cognitive sciences that has moved away from innate universals and clear-cut modules (e.g., Fodor 1975; Fodor 1983; Caramazza & Mahon 2006; Caramazza & Shelton 1998; Pinker 1994) and towards understanding humans’ minds – including language – as highly plastic, embodied and situated entities that vary between individuals and cultural groups, are grounded in action, and embedded in situated contexts in the physical world (e.g., Barsalou 1999; Clark & Chalmers 1998; Barsalou 2008; Tomasello 1999; Tomasello 2008; Tomasello 2014; Hutchins 1995). Human language is thus not just biologically determined, but also subject to a wide variety of pressures. But what evidence is there that this variation is – at least in part – the result of some form of adaptation and not just due to random drift?

Many of the pressures exerted on language are internal, and either related to symbolic communication itself, or the cognition of its users. For example, expressivity and simplicity have shaped language over the course of cultural evolution to facilitate learning, transmission and processing (see Zipf 1948; Kirby 2017; Christiansen & Chater 2008; Christiansen & Chater 2016; Kanwal et al. 2017). This can, for instance, be seen in category systems (Carr et al. 2020), such as those used for classifying kinship (with categories like *aunt*, *cousin* etc.), which across all languages show an optimal balance of informativeness and learnability considering the space of theoretically possible systems (Kemp & Regier 2012). Similar findings have been reported for other semantic domains, such as colour terms, spatial adpositions, numerals, and quantifiers (Carstensen et al. 2015; Regier, Kemp & Kay 2015; Xu, Regier & Malt 2016; Steinert-Threlkeld & Szymanik 2019; Xu, Liu & Regier 2020). Similarly, cognitive biases present in language learners, such as a bias for simplicity can become amplified over iterated learning, that is the transmission of linguistic forms from one generation of learners to the next (Kirby, Griffiths & Smith 2014). This domain-*general* bias for simplicity that is related to multiple pressures for learnability, has been observed

to shape *specific* domains across languages, such as statistical tendencies in word order patterns across languages (Culbertson & Kirby 2016).

However, there are also external pressures on the linguistic system. If we look at word order, there is quite a lot of variation across languages, but this variation does not appear to be random: For instance, it has been found that there is an isomorphism in how languages order collections of modifiers, e.g., *those three green apples*. In both pre-nominal languages (e.g., English) and post-nominal languages (e.g., Thai), speakers tend to prefer demonstratives to be further away from the noun than numerals, followed by adjectives. Artificial language learning experiments have revealed that this linguistic tendency seems to be explained by a cross-culturally robust cognitive bias for hierarchical structure (Culbertson & Adger 2014; Martin et al. 2019). However, the question remains: where does this bias come from in the first place? Is it simply innate, or could it reflect the logical structure of the world? Recent work by Culbertson et al. (2020) shows that these preferences exist not only in languages, but also improvised gestures. The authors argue that this reflects statistical properties of the world. Adjectives, i.e. properties like colour, would simply be more related to objects than say, numerals. They tested this with a cross-linguistic corpus study using information-theoretic measures and found that the strength of association between different modifiers predicted this word order pattern (placing the demonstrative farthest away and the adjective closest to the noun). In other words, this study provides strong evidence that even internal cognitive biases can, in turn, result from our interaction with the world, in this case reflecting statistical properties of the objects we encounter in our daily lives.

There are more examples of linguistic adaptation from various other domains, including relationships between linguistic features and social, physical, physiological, anatomic, genetic, and technological variables, some of which were already briefly mentioned in Paper 1 above.

The social domain includes the relationship between morphological complexity and population size, which suggests that larger communities display a lower complexity on various measures (Lupyan & Dale 2010). There are also a number of studies that have looked at economic variables and behaviour, for instance relating the marking of future tense to propensity for spending or saving (Roberts, Winters & Chen 2015).

Physical effects of the environment have been suggested to have an impact on the lexicon or sound systems. For example Regier et al. (2016) found that languages spoken in warmer climates are more likely to use the same word for ice and snow, which they explain with language adapting lexically to allow for efficient communication about the local environment. More recent studies looking at colour terms have similarly suggested that colour lexicons result from a combination of perceptual constraints (Gong et al.

2019) and communicative need (Gibson et al. 2017; Zaslavsky et al. 2018). Naturally, the importance of distinguishing specific colours varies in relevance with respect to individual or cultural differences. For example, the police might be interested in a green car that was stolen, while a customer choosing the colour of their car might be deciding between mint green and bamboo green. It appears to be the case that similar to such individual registers that might vary between speakers of the same language (say between artists, and non-artists), the colour term inventories of languages more generally seem to reflect the communicative needs of their speakers. Gibson et al (2017) compared three groups, the Tsimane, an Amazonian hunter-gatherer society, Bolivian-Spanish speakers, and American English speakers. They found that for all groups warm (rather than cool) colours were more ‘useful’ to talk about daily objects and scenes, which is thus reflected in the lexicon. At the same time, the Tsimane were less likely to use colour terms for familiar objects, which the authors interpret as an effect of industrialization promoting colour-naming efficiency. Lastly, in the domain of colour, even physiological adaptations have been suggested, such as weakened tendency to lexicalise ‘blue’ colour terms associated with phototoxic effects in high UV-A environments (Lindsey & Brown 2002) and the prevalence of daltonism near the equator and in lower latitudes (Brown & Lindsey 2004; cf. Regier & Kay 2004).

Other climatic effects have been proposed to influence sound systems, with languages spoken in more humid climates being more likely to exhibit tone as a feature (Everett, Blasi & Roberts 2015), fewer vowels (Everett 2017) and other correlations related to altitude (Everett 2013b), vegetation or ambient temperature (Maddieson & Coupé 2015).

Other research has found adaptations that seem to be more related to biological ‘environments’. For example, Dediu et al (2017) propose that the relative rarity of clicks in most phoneme inventories might be the result of a small anatomical biases in vocal tract shapes that became amplified over cultural evolution, and Dediu and Ladd (2007) propose a similar mechanism based on a correlation between two genes and the distribution of linguistic tone.

Finally, language appears to manifest itself in a wide variety of modalities, such as the signed, spoken, whistled or visual (Perniss 2018), and seems to easily adapt to new technologies that pose very particular demands, such as writing (Morin, Kelly & Winters 2018), written chats and emojis (Lupyan & Dale 2016) or even completely novel communication channels including scratch pads, slide whistles and other signalling spaces that have been used to study symbol grounding in the lab (see Tamariz 2017 as well as the following section and later chapters for more examples of such experiments).

Considering these findings, it appears that languages differ not just due to random drift. There is now evidence that this diversification due to adaptation has even shaped the linguistic family tree, which

can be detected in phylogenetic signals that account for social, physical or ecological factors (Bentz et al. 2018).

But how much can we trust correlational findings? Large datasets are naturally prone to spurious correlations. For example, Roberts and Winters (2013), report several such effects, such as a strong correlation between linguistic diversity in a given country and the number of accidents (more road fatalities correlate with higher diversity). However, they point out that from such data, we cannot simply draw causal conclusions of the kind ‘more diversity means more miscommunications, means more traffic accidents’. Linguistic features are cultural datapoints and as such confounded by variables such as historical descent and geographical proximity. The autocorrelation of cultural datapoints is also known as ‘Galton’s problem’ and is often ignored in studies that are frequently conducted by non-linguists that seem to be unfamiliar with the analysis of cultural data (e.g., Chen 2013; Kim, Kim & Zhou 2017; Feldmann 2019), which can lead to overconfidence in false positives. However, the literature on linguistic adaptation is gradually changing, and many results are now being critically assessed and newer studies usually distinguish causal from correlational effects. For example, Hua et al. (2019) report a climatic effect on the global distribution of language diversity that coincides with biodiversity. However, the latter is an incidental effect, due to climate influencing biodiversity, which has no causal link to language.

In this context, S. G. Roberts (2018) has therefore argued for a more rigorous, robust and incremental approach to studying linguistic adaptation. As one of the co-authors on the aforementioned study on humidity and tone (Everett et al. 2015) that has frequently been criticised (see Everett, Blasí & Roberts 2016a), he himself points out problems with relying purely on cross-sectional data. Rather than focusing on finding statistical patterns, research in linguistic adaptation should be hypothesis-driven and rely on concrete causal reasoning. This is especially necessary since most phenomena do not simply involve two variables. For example, the relationship between morphological complexity and population size has been associated with a large number of factors (e.g., number of second learners, see Bentz & Winter 2013) can be represented in a causal graph with 19 hypotheses that have been formulated in the literature (see graph in Roberts et al. 2020: 13). Representing all possible variables and their relationships explicitly, then allows testing these hypotheses to uncover the actual mechanism. Some studies on the distribution of speech sounds have already incorporated this approach that integrates large-scale observational data with experiments. For example, Dediu et al.’s (2014, 2019) work combines biomechanical modelling to simulate the production of speech sounds with human subject experiments and large-scale phylogenetic analyses of the distribution of sounds across languages. Dediu et al. (2019) were able to show that changes in diet after the Neolithic revolution resulted in anatomic changes, which in turn made it easier to produce labiodental

sounds (like [f] and [v]). This bias then became amplified through cultural evolution and shaped the sound systems after the Neolithic. To convincingly show that more humid climate can give an advantage to the emergence of tone in a language, it is necessary to assess a similar chain of causal links. To mechanistically show how ambient humidity affects the production of tones one could for example combine experiments that measure ease of production in environments with different humidity levels and then use an iterated learning study with an artificial language, to test whether there is a bias that can actually be amplified through cultural transmission (see Roberts 2018 for a detailed proposal).

In the following section, I will give an overview about this experimental approach that has been used to study evolutionary mechanisms shaping language with human participants, which, in my view, provides ideal paradigms for supplementing correlational data with experiments that can test explicit causal hypotheses. I should also note here that the issue of correlation and causation as well as the CHIELD database that helps representing hypotheses from the literature in causal graphs, will be discussed in some more detail in [Paper 3](#), where Roberts' (2018) robust and causal approach is applied to the adaptation of spatial language.

1.5 An experimental evolutionary approach for investigating linguistic adaptation

As discussed in [Paper 1](#), language evolution research is a diverse field that is both interested in the biological capacity for language as well as the cultural evolutionary mechanisms that have and continue to shape linguistic structure, both of which interact to some extent (see, e.g., Fitch 2017; Kirby 2017). In this thesis, I will mainly focus on the cultural side, that is, how language is shaped as it is learned, used for communication, and transmitted by its users. In recent years there have been some methodological innovations in the study of cultural language evolution, which I believe are highly relevant to the study of linguistic adaptation. Most notably, the last decade saw the beginning of an experimental approach to study how communicative factors and cognitive biases (such as the ones briefly discussed above) can give rise to different kinds of linguistic structure. The first studies focused on fundamental design features, such as compositionality and combinatoriality and the complementary effects of communication and transmission on these features (Tamariz 2017 for reviews). The two main strands are the Iterated Learning Model (Kirby, Griffiths & Smith 2014) and the Experimental Semiotics paradigm (Galantucci, Garrod & Roberts 2012). Iterated Learning studies have shown that transmitting unstructured artificial languages from one generation of learners to the next creates a cognitive bottleneck through which a bias for regular and learnable forms becomes amplified and manifests itself in increasingly compositional languages that emerge over

generations (Kirby, Cornish & Smith 2008). By contrast, Experimental Semiotics studies started out by focusing on interaction and testing how dyads solve referential communication games if they are deprived of the ability to use their language (see, e.g., Galantucci 2005). Both approaches originally started with computational simulations (e.g., Smith, Kirby & Brighton 2003; Kirby & Hurford 2002; Steels 1995), but have since been successfully replicated and expanded upon with human subject experiments, which allows drawing conclusions about actual human behaviour and cognition (Galantucci, Garrod & Roberts 2012; Kirby, Griffiths & Smith 2014).

As far as linguistic adaptation is concerned, a few studies from this area have begun to also consider the role of language-external variables, such as the referential environment or communicative factors in these types of experiments. For example, a study by Silvey and colleagues (2015) found that manipulating the relevance of the meaning dimensions from Kirby et al.'s (2008) original iterated learning study, could result in underspecification of dimensions that were less salient (e.g., if shape or colour was backgrounded at the learning stage, before the language was transmitted). In a similar iterated learning experiment Tinits et al. (2017) found that participants were more likely to gradually over-specify a non-obligatory colour marker in specific communicative contexts, where overspecifying this dimension could reduce cognitive effort, which could help explain, why some languages evolve seemingly redundant markers (e.g., for animacy). Similarly, Winters et al. (2015) found that depending on which meaning dimensions (such as shape or colour) are present in the referential context of a dyadic referential communication game, languages will emerge that are either holistic, fully systematic, or underspecified with respect to these dimensions. Additionally, a related study showed that predictability in the referential context shapes the degree of signal autonomy, i.e. languages used to communicate about highly predictable contexts depend on this context to a larger extent in order to reduce semantic uncertainty (Winters, Kirby & Smith 2018). Other studies looking at artificial communication systems emerging in interactions have found that word order, likewise, can be influenced by structural properties of the referential context that is communicated about, for example in relation to event types (Schouwstra & de Swart 2014; Christensen et al. 2016). In other words, there is some evidence that artificial languages emerging in the lab are just as sensitive to the communicative and environmental context in which they are used as natural languages.

In this thesis, I therefore adopt this experimental evolutionary approach to further investigate linguistic adaptation in two steps: Firstly, there is now abundant experimental evidence that systematic linguistic structure can emerge due to constraints from interaction and transmission. I want to expand on this by studying to what extent factors relating to the communicative or referential environments can mediate this process, i.e. if manipulating these environmental variables can make specific systematic

categories or sign types more likely to emerge in communicative interaction. Such systematic structure is fundamental to communication systems, since it allows abstracting away from item-based individual signals and creating a *system* of forms enabling combinatorial structure (such as in phonology) and compositional structure (such as in morphosyntax, see Hockett 1960; Dingemanse et al. 2015).

This would show that, at a fundamental level, linguistic structure is sensitive to input from the environment, which can lead to diversification of communicative codes from the very start as they emerge. Secondly, I apply this approach to a more complex real-world phenomenon to show that experiments can be used to test causal links in otherwise messy and complex relationships between linguistic and environmental variables. In the following section, I will briefly outline the projects that attempt to actualise both of these steps.

1.6 Roadmap

In this chapter, I have laid out the main framework and method – using laboratory experiments to study how the external environment can shape the diversity of linguistic structure. This thesis is a cumulative effort that includes various research projects in the area of language evolution and linguistic adaptation that all attempt to address the following general hypothesis:

Emerging semiotic communication systems that evolve interactions and transmission are ecologically biased. Such systems will reflect cognitive strategies that have been adaptive to environmental factors

I will first focus on the more fundamental property of systematic structure in language, i.e. the property allowing languages to mark statistical relationships between meanings and forms, which enables productivity and facilitates learning. Chapter 2 presents a study (Paper 2, Nölle et al. 2018b) exploring the emergence of systematic structure in two ‘silent gesture experiments’ (see, e.g., Goldin-Meadow et al. 2008; Motamedi et al. 2019; Schouwstra, Swart & Thompson 2019). Participants had to solve a communicative task where they had to ground a gestural communication system in order to communicate about drawings. We found that manipulating aspects of the referential environment as well as communicative factors affected how likely they were to develop systematic categories and express these with systematic gestures. ‘Systematic’ here refers to conventionalised recurrent gestures that were applied to more than one referent, thus creating abstract categorical forms that allowed for productive compositionality (e.g., marking two referents as *female* using the same gesture).

Additionally, these environmental variables affected which referential dimensions would be systematized and how stable these systematic markers remained over time.

The second part of this thesis will focus on a more concrete case of linguistic variation, namely spatial language, which has been proposed to vary with environmental factors such as topography or subsistence style (e.g., Palmer et al. 2017). In chapter 3, I will first give a short introduction to spatial language and its complex relation to culture, cognition and the physical environment. I will then present a paper (Paper 3, Nölle & Spranger, accepted) that argues that in order to understand these complex relationships and the origins of different spatial referencing strategies, we have to combine traditional comparative fieldwork with evolutionary experiments that can test specific relationships, such as between spatial Frames of Reference and the presence of certain topographic features.

In the remaining chapters, I show a series of experiments used to test the relationship between spatial language and topography directly in the lab. To this extent I introduce virtual reality (VR) experiments as a new method to both research in language evolution and spatial language. Chapter 4 presents two VR experiments testing how speakers whose language relies predominantly on an egocentric reference frame, solve the same communicative task in large-scale environments with different environmental affordances. The results indicate that the environment motivates differences in how they solve the task and that participants can flexibly switch between strategies available to their language in response to topographic factors.

Chapter 5 presents a third VR experiment, where participants could not use their native language, to test whether different strategies would also emerge across environments in a novel communication system as well as how conventionalized systems would fare if participants switch between different environments. I also report a novel way to use spatial movement data in order to analyse the meaning of novel signals that emerge during the experiment.

Chapter 6 provides a general discussion of the findings from all three experiments, some limitations as well as possible directions for follow-ups and future experiments.

Chapter 7 is a short conclusion to this thesis.

2 Linguistic adaptation in the Laboratory

2.1 Preface to Paper 2

Paper 2 (Nölle et al. 2018b) began as a collaboration in 2016, before I began my PhD. Data for experiment 1 was collected in Spring and Summer of 2016 while I worked as a research assistant in a project titled *Contextual Factors Shaping the Emergence of Communication Systems* funded by the Interacting Minds Centre at Aarhus University in Denmark. The second experiment was conducted in 2017, during the first year of my PhD at Edinburgh. I travelled to Aarhus in February 2017 to collect the data. All coding of video materials, data analysis and writing for both experiments took place during my PhD.

The basic idea behind the paper – attempting to disentangle holistic iconic signs and systematic signs emerging in the laboratory – builds on an earlier idea from Kristian Tylén and evolved into the eventual project through discussion with me and Marlene Staib. The experimental design was conceived by all three of us. I created the stimulus materials and Marlene Staib programmed the interface with input from myself. Both of us collected the data in the laboratory. Riccardo Fusaroli gave advice on the statistical models used in the paper. The analysis was performed by Marlene Staib and Riccardo Fusaroli and revised by me. The first draft of the paper was written by me and Marlene Staib. I revised the paper for the final version. All authors provided comments.

This version of the paper constitutes the last version sent to the journal. The final version published in 2018 in *Cognition* contains minor changes made to the text at the proof stage. This final version of the paper is copyrighted by the publisher and can be accessed at <https://doi.org/10.1016/j.cognition.2018.08.014>. Note that in the thesis version of this paper, figure and table numbers have been altered to be consistent with the rest of the document. All references cited in Paper 2 can be found at the end of this thesis.

2.2 Paper 2:

The emergence of systematicity: how environmental and communicative factors shape a novel communication system

Jonas Nölle, Marlene Staib, Riccardo Fusaroli, and Kristian Tylén

Abstract

Where does linguistic structure come from? We suggest that systematicity in language evolves adaptively in response to environmental and contextual affordances associated with the practice of communication itself. In two experiments, we used a silent gesture referential game paradigm to investigate environmental and social factors promoting the propagation of systematicity in a novel communication system. We found that structure in the emerging communication systems evolve contingent on structural properties of the environment. More specifically, interlocutors spontaneously relied on structural features of the referent stimuli they communicated about to motivate systematic aspects of the evolving communication system even when idiosyncratic iconic strategies were equally afforded. Furthermore, we found systematicity to be promoted by the nature of the referent environment. When the referent environment was open and unstable, analytic systematic strategies were more likely to emerge compared to stimulus environments with a closed set of referents. Lastly, we found that displacement of communication promoted systematicity. That is, when interlocutors had to communicate about items not immediately present in the moment of communication, they were more likely to evolve systematic solutions, supposedly due to working memory advantages. Together, our findings provide experimental evidence for the idea that linguistic structure evolves adaptively from contextually situated language use.

1. Introduction

Systematicity permeates language at all levels. Most languages, for instance, show consistent constituent orders (e.g. Subject-Object-Verb vs. Subject-Verb-Object), compositionality (e.g. systematic marking of tense, gender, case, number etc.), semantic roles or even systematic sound-meaning mappings as in phonestheme clusters such as *glimmer*, *glare*, *glisten* (Bergen 2004; Monaghan et al. 2014). Systematicity thus refers to statistical relationships between forms that relate in their meanings, ultimately constituting ‘categories’ on the form side. But where does systematicity come from? What are the cognitive and communicative factors that promote the persuasive propagation of systematicity across almost all aspects of linguistic structure?

Controversies exist in the language sciences concerning the foundations of linguistic structure. Some theories favour biological explanations grounding linguistic structure in innate genetic code (Chomsky 1986; Hauser, Chomsky & Fitch 2002; Pinker 1994). Another prevalent suggestion is that systematicity emerges in response to internal, cognitive biases that get amplified through transmission and learning in processes of cultural evolution (Brighton 2002; Smith, Kirby & Brighton 2003; Christiansen & Chater 2008; Kirby, Cornish & Smith 2008; Kirby, Griffiths & Smith 2014; Kirby et al. 2015). Using the ‘iterated learning paradigm’, Kirby and colleagues have investigated how sign systems become increasingly ordered, compressible and easier to learn and process, as they are transmitted from one generation to another, due to subtle unintended distortions as signs pass through cognitive bottlenecks and biases of language learners (Kirby et al. 2008). However, it is unclear where this human propensity for systematicity comes from and how specific features are selected and become expressed in systematic categories.

In this paper, we explore the circumstances under which systematic strategies evolve in communicative interaction when referents can also potentially be disambiguated using idiosyncratic, holistic signs. By “circumstances”, we mean factors that pertain to aspects of the referent environment and communicative situation. The idea that linguistic structure is promoted by a number of external, contextual factors has recently been coined the Linguistic Niche Hypothesis (Lupyan & Dale 2010).

In particular, we suggest that systematicity in language emerges adaptively in response to environmental and social factors associated with the situated interactive practice of communication itself (rather than individual learning, Tylén et al. 2013). That is, linguistic structure is motivated by, and evolves contingent on, structural properties of the physical and social environment. Recent studies provide cross-sectional/correlational evidence for the idea that linguistic structure is contingent on environmental factors, thus suggesting that languages evolve adaptively to meet ecological affordances. Examples include climate influencing the lexicon (Lindsey & Brown 2002; Regier et al. 2016) or sound systems (Everett, Blasi & Roberts 2015; Everett, Blasi & Roberts 2016b), subtle genetic or anatomic biases guiding variation (Dediu, Janssen & Moisik 2017; Dediu & Ladd 2007), as well as social factors such as number of speakers that has been found to affect morphological complexity (Lupyan & Dale 2010; Lupyan & Dale 2016).

In this study, we take an experimental approach to address the question whether there are specific environmental circumstances under which systematic categories and signs are more adaptive and thus more likely to evolve in competition with non-systematic strategies that might also provide viable solutions. We first define what distinguishes systematic signs from idiosyncratic signs and then individuate three complementary environmental factors hypothesized to promote systematicity in an evolving

communication system: i) inherent structure of the referent scenes, ii) instability/openness of the referent environment, and iii) displacement of the communicative environment from the referential scene

1.1 Functional pressures for systematicity, iconicity, and arbitrariness

Systematicity contrasts with other referential principles such as iconicity and arbitrariness that describe the relation between single signs and their meanings. De Saussure (1959) famously argued the defining trait of language to be ‘arbitrariness of the linguistic sign’ and thus inaugurated a widely endorsed linguistic tradition relegating non-arbitrariness to the role of a rare and peculiar phenomenon to be mostly ignored. However, recent work has highlighted subtle motivations underlying linguistic structure (Kirby, Dowman & Griffiths 2007; Lupyan & Dale 2016). Examples include studies of ideophones, sound symbolism and systematicity (Dingemanse 2012; Dingemanse et al. 2015; Monaghan, Mattock & Walker 2012). A general tendency in this literature has been to subsume iconicity and systematicity under ‘non-arbitrary forms’ as opposed to ‘arbitrary forms’ (e.g., Monaghan et al. 2014). However, iconicity and systematicity could in fact be argued to be the expression of diverse adaptive pressures (Dingemanse et al. 2015).

Iconic signs are motivated in that there is a resemblance between their form and meaning (Peirce 1931). Iconicity has thus been suggested to play a prominent role in the grounding of communication systems as mappings between form and embodied experience on both phylo- and ontogenetic time scales (Fay, Ellison & Garrod 2014; Harnad 1990; Perniss, Thompson & Vigliocco 2010). In language acquisition, iconic sound-referent mappings seem to facilitate early word-learning (Imai & Kita 2014; Imai et al. 2008; Perlman et al. 2017; Perry, Perlman & Lupyan 2015) as well as novel word learning in adults (Bergen 2004; Lockwood, Dingemanse & Hagoort 2016). In addition, experimental lab studies of emergent communication systems indicate iconicity as one of the main strategies employed whenever signs are grounded from scratch in interaction (Fay et al. 2014; Perlman, Dale & Lupyan 2015; Perlman & Lupyan 2018). For example, in studies where participants invent new communication systems using a graphical medium, iconicity serves as a starting point for bootstrapping communication, after which signs become gradually simpler and more symbolic (Caldwell & Smith 2012; Garrod et al. 2007).

However, if iconicity provides such efficient means to ground a communication system, why do natural languages only display rather subtle aspects of iconicity? And why do we generally observe iconicity to decay over time in favour of more systematic and arbitrary mappings (Garrod et al. 2007; Little, Perlman & Eryilmaz 2017)? As evidenced in, for instance, emerging sign languages (Senghas 2004; de Vos & Pfau 2015) home signs (Haviland 2013; Mylander & Goldin-Meadow 1991) and semiotic experimental studies (Galantucci, Garrod, & Roberts 2012; Tamariz 2017), signs systems often set off as a set of

individual, idiosyncratic mappings to referents (Deacon 1998). Over time, the sign repertoire evolves and regularities among and relations between individual signs emerge, which eventually give the repertoire properties of a system. As relations among a set of signs stabilize, they provide an alternative means to ground new signs: the meaning of a sign is thus not only constituted by the concrete mapping to a referent, but also by its more abstract relations to other signs in the system. Resulting statistical regularities have been suggested to shift the mnemonic strategy of learners to rely increasingly on the relations among signs (Deacon 1998) and allow them to quickly categorise newly encountered signs, generalize them and incorporate them into the wider system. Indeed, studies have shown that language-specific statistical differences in word forms (e.g., verb vs. noun morphology) serve as cues that assist category learning (Cassidy & Kelly 1991; see Dingemanse et al. 2015 for a review; Monaghan, Chater & Christiansen 2005; Monaghan, Christiansen & Chater 2007). Consequently, signs gradually lose their motivated connection to referents and become increasingly systematic as they come to increasingly depend on their interrelations internally within the communication system.

Thus, while iconicity and systematicity have often been treated as an expression of the same basic pressure of ‘motivation’ compared to arbitrariness (e.g., Monaghan et al. 2014), they might be better conceived as orthogonal pressures related to different adaptive pressures (see also Dingemanse et al. 2015; Nielsen 2016).

1.2 Outline of the study

The current study was designed to address the circumstances under which systematicity evolves in a novel sign system even if idiosyncratic signs are equally afforded. In many situations both strategies would apply: For instance, if one were to point out a specific person among a crowd of people, one could rely on a single discriminating idiosyncratic trait such as “the person with the funny-looking beard” or “the individual with shiny red shoes”: However, one could also disambiguate the referent by pointing to a specific combination of more general traits, like gender and job category, as in the example “the female doctor” (as opposed to male doctors). When grounding a novel communication system, what are the conditions that promote the latter systematic (analytic) strategy in contrast to a simple idiosyncratic (iconic, holistic) strategy? That is, when is systematicity more adaptive than encoding single features of referents in one-to-one form-meaning mappings?

Building on the general assumptions of the Linguistic Niche Hypothesis suggesting that systematicity evolves adaptively in response to particular ecological and social affordances, this study sets out to test three complementary hypotheses:

1) Systematicity in communication systems is motivated by regularities in the environment. When new signs evolve under pressure for social coordination and communication about objects and events in the environment, salient relations among those referents provide a semiotic resource, motivating systematic structure among emergent sign forms (Christensen, Fusaroli & Tylén 2016; Lupyan & Dale 2016; Tinitis, Nölle & Hartmann 2017; Tylén et al. 2013; Winters, Kirby & Smith 2015). Following such predictions, people will be more inclined to selectively systematize those dimensions of their communication system that also appear more structured in the referential environment (Christensen, Fusaroli & Tylén 2016; Nölle 2015). Returning to the example above, marking the gender of a specific referent would be more adaptive in an environment where the trait is discriminative and thus helps disambiguating between competing referents, that is, in a crowd that consists of both females and males.

2) Systematicity is afforded by ever-changing or unstable environments. Here, we focus on a particular kind of instability, open-endedness. Grounding new signs involves considerable coordination and processing costs for the communicating parties. In “open” environments with novel incoming or changing referents, systematic compositional communication systems provide coordination advantages due to their productivity and thus allow new signs to be scaffolded by, or even constructed from, already existing signs (Carr et al. 2017; Selten & Warglien 2007). Open environments should therefore motivate more attention to abstract regularities that generalize across tokens in the referential context. In other words, we predict systematicity to evolve in response to dynamic environments where novel referents are continuously introduced into the communicative context.

3) Lastly, systematicity is afforded in contexts of displaced communication (when communicating about absent referents). Most studies of emergent communication systems take place in situations where referents are co-present to the interlocutors (e.g., Fay, Arbib & Garrod 2013; Garrod et al. 2007; Scott-Phillips, Kirby & Ritchie 2009). In these situations, less load is put on the working memory of the interlocutors as they can continuously access the referents to incrementally disambiguate the referent target. The need to communicate about episodic content not present in the moment of communication itself provides external pressure on working memory, e.g., by displacing the referential context from the communicative context. This increases the processing load since idiosyncratic features of referents cannot be directly and continuously attended to in the environment but have to be kept in memory. We thus predict more systematicity in contexts of displaced communication as it drastically reduces the working memory load by “chunking” the idiosyncratic properties of many possible referents into more general and economic categories (Christiansen & Chater 2016).

In sum, we hypothesize that systematicity is motivated by structural properties of the physical and social environment, and in particular i) that regularities in sign systems reflect regularities in the environment through the principle of functional adaptation, ii) systematicity subserves productivity and is therefore more likely to emerge in an unstable, “open” environment with continuous introduction of new referents, and iii) systematicity will increase in contexts of displaced communication (when communicating about absent references) due to a working memory bottleneck. Furthermore, we predict systematicity to emerge over time during interaction in response to the above factors, and that openness of the environment and displacement are likely to interact with time yielding gradually higher levels of systematicity in displaced communication unfolding in open environments.

In order to investigate how the factors described above modulate the relative attraction to competing strategies in the course of stabilizing a novel communication system, we constructed stimuli that would lend themselves to multiple referential solutions. In other words, a target referent in the stimulus set could both be disambiguated using an idiosyncratic iconic or a systematic strategy. In this context, an idiosyncratic strategy would be to use an iconic sign to refer to the referent by attracting attention to a maximally discriminable, unique trait (e.g., a detail that is not shared with any other stimulus item). In contrast, a systematic strategy would entail using combinatorial, analytic forms to draw attention to certain combinations of traits shared across stimulus items and thereby disambiguate the target referent.

We used the “silent gesture” paradigm (Christensen, Fusaroli & Tylén 2016; Fay, Arbib & Garrod 2013; Schouwstra & de Swart 2014) in which dyads have to solve a referential task by inventing a novel gestural communication system. Experiment 1 tests hypotheses 1 and 2, while experiment 2 tests hypotheses 2 and 3.

2. Experiment 1: Systematicity is motivated by structural properties of the environment

2.1. Methods

2.1.1. Participants

27 pairs of participants (9 male, 8 female, 10 mixed, mean age 22.9, SD 5.3) took part in the experiment in return for monetary compensation. Participants were recruited among students at Aarhus University, Denmark. Some of the pairs knew each other in advance. All were native speakers of Danish.

2.1.2. Materials

24 cartoon images of different human characters were created using the freely available online-software bitstrips (Blackstock & Brown 2009) and modified for the purpose of the study with the open-source editor Inkscape (The Inkscape Team 2003). Each character was designed to be individually identifiable from among the full set by either a unique feature (e.g., necklace, glasses, beard etc.) or as a member of three types of categories – PROFESSION (e.g., cleric, construction worker, chef), GENDER (male/female) and PET OWNERSHIP (yes/no). These traits were distributed evenly across the full set in such a way that there was one character per combination of categories (e.g., female soldier with a pet; see Figure 2-1 for examples and supplementary material for the full set), leading to a total of four characters (male/female, with/without pet) per profession. In the experimental task, this equally afforded two competing solutions: Using silent gesture, participants could either refer to unique features of the complex stimuli by using iconic gestures (e.g., gesturing glasses) or abstract away from these concrete features and gesture combinations of general categories in a systematic way to disambiguate referents (e.g., male chef without pet). Pseudo-random subsets of 14 stimuli were created for presentation, keeping the following constraints: Each subset included two characters from each of the six job categories, and one additional pair from one of the categories. One of the binary traits (PET, GENDER) was chosen to be the more functionally adaptive trait in the set, that is, it was distributed evenly over the selected meaning space (giving a 50-50 split) and therefore had a higher discriminative value, i.e., helped individuate characters within the same job category (see Figure 2-2). The other, less functionally adaptive trait was distributed 10-4 within the set and thus had a lower discriminative value, both overall and within categories. To control for markedness (e.g., ‘female’ might be more (conceptually) marked than ‘male’, and ‘with pet’ more marked than ‘without pet’), the distribution of the less functional trait was counterbalanced across pairs/sets (e.g., sets with 10 males and 4 females vs. sets with 10 females and 4 males).



Figure 2-1: Example Stimuli. Members of the same category share colours, which cannot easily be represented by iconic gestures, but make the categories more salient.

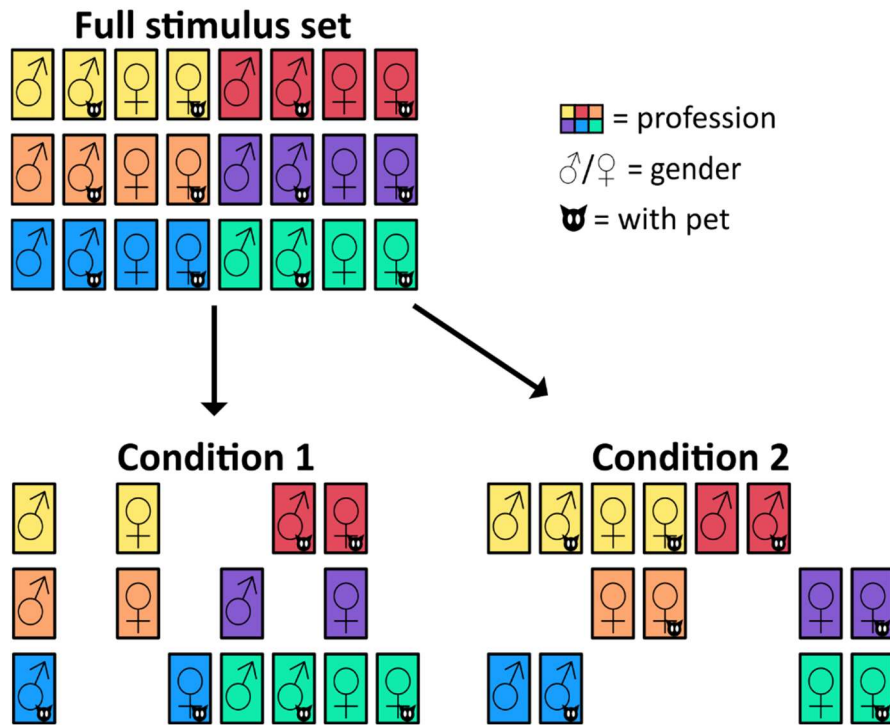


Figure 2-2: Example trial sets. In condition 1 on the left, the gender trait divides the meaning space in equally sized subsets and is the best discriminating feature within job categories. In

condition 2 on the right, the pet trait is more functionally adaptive. Both traits do however occur in both conditions.

2.1.3. Design and procedure

The study employed a two-by-two factorial design with functional adaptivity (high/low) as the first factor and referent environment (open/closed) as the second factor. Functional adaptivity was manipulated within pairs, as one trait (PET/GENDER) was always chosen to be more functionally adaptive than the other within the stimulus set a pair received. The distribution of traits (GENDER more discriminative/PET more discriminative) was counterbalanced across pairs. In addition, “markedness” of the traits was also counterbalanced. For example, in a set in which gender was more functionally adaptive, a large number of stimuli depicted with pet would make the trait PET still more salient compared to a set where most stimuli are depicted without pet.

Pairs were assigned randomly to one of two referential environment conditions: In the ‘closed’ condition, participants were, on any given trial, presented with one target and 13 foils (i.e. non-target images) drawn from the same set of 14 stimulus items. This was contrasted with the ‘open’ condition, in which the same targets were still drawn from a stable set of 14 items while foils were sampled from the full set of images and kept changing on every trial, although retaining the distributional properties of the first set. Therefore, the functional adaptivity of traits nevertheless remained stable. This allowed simulating a more ‘open’ environment, where new referents are constantly introduced, while keeping the demands on communication and memory constant across conditions.

Upon entering the lab, participants were informed that this was a study about the evolution of novel sign languages, and that for this purpose they had to use gestures to communicate about visual stimuli, without talking or making any communicative sounds. They then gave their informed written consent, including permission to video-record their interactions. Further written instructions explaining the details of turn-taking and feedback were given on screen controlled by the stimulus presentation software PsychoPy2 (Peirce 2007). Partners were seated across from each other at a table, each with a 22-inch (1920 x 1080) monitor set up to one side (Figure 2-3). In each trial, one participant was assigned to be the gesturer, who had to communicate a given stimulus to her partner, and the other to be the guesser, who had to identify that stimulus from the set of 14 stimuli. All 14 stimuli of a set were visible to both the gesturer and the guesser at all times, with a red frame indicating to the gesturer which stimulus she had to communicate. Once the guesser had selected one of the referents from the set by clicking on it with the mouse, auditory feedback was played to indicate success or failure, before moving on to the next trial. Roles were switched after every trial. Even though participants were not explicitly instructed to do so, turn-taking

within trials was allowed, meaning that the guesser could ask “questions” (in gestures) or signal understanding. The experiment consisted of four rounds, during each of which each stimulus was communicated once by each pair member (in randomized order), yielding a total of 112 trials per pair. Interactions were recorded with a GoPro Hero 4 camera.

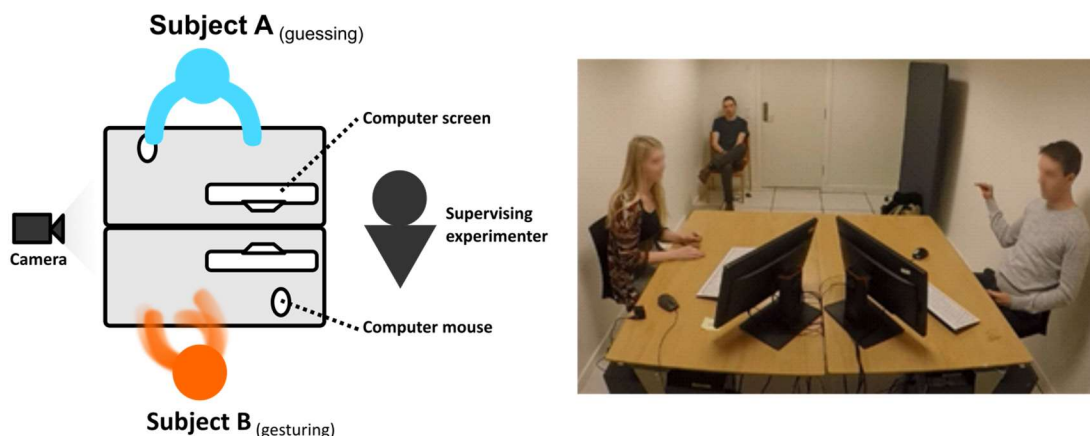


Figure 2-3: Experimental setup.

2.1.4 Analysis

Coding scheme: Gesture strings were analysed on the dyadic level, that is, any gestures produced by either the gesturer or the guesser during a trial were considered to be part of the gesture string for that trial. Individual gestures were classified as systematic by a human coder whenever the same gesture was used to communicate at least two different stimulus items. For example, if the gesturer produces a “stroking” gesture to refer to a pet (e.g., the dog next to a soldier), and the same gesture was later used to refer to another pet (e.g., a cat next to a priest), then the gesture is counted as systematic. However, if a new gesture, for instance “paw licking”, was used to refer to second stimulus (the cat), it was not counted as systematic, but as idiosyncratic/iconic. The classification only applied looking backward, i.e., from the second referent onwards (and not to any gestures that would only later be used for another referent and thus get systematized). Systematic gestures were further categorized into job, pet and gender gestures – thus allowing for a maximum of 3 (and a minimum of 0) instances of systematicity on every trial. The two first authors and one independent coder blind to the hypotheses each coded a third of the data, with 7 percent overlap between each coder pairing (21 percent overlap overall).

Analysis: Intercoder reliability was assessed using percent agreement and unweighted Cohen’s Kappa (Cohen 1960) using the package ‘irr’ for R (Gamer et al. 2012). Percent accuracy (for guessing the correct stimulus on each trial) was calculated for each condition to confirm the emergence of an efficient communication system, and the development of accuracy over time was estimated using a mixed effects logistic

regression model (with random intercepts for subject/pair/condition, coder and stimulus). The number of gestures used (by either the gesturer or the guesser) on each trial was counted and their development over time was analysed analogously to accuracy using a linear mixed effects regression model. The development of the number of gestures per trial was used as a measure of simplification and conventionalization, as it is typically found in experimental studies on emergent communication systems (Caldwell & Smith 2012).

In order to test whether hypotheses 1 and 2 were supported by our experimental data, we built a logistic mixed-effects regression model with systematicity (per each feature) as binomial outcome variable and referent environment (open/closed), functional adaptivity (low/high), (z-standardized) trial number (1-112) and markedness (trait marked/unmarked) as predictors. Initially, the maximum random effects structure justified by the design was chosen (Barr et al. 2013), excluding random slopes for interactions to ensure convergence and avoid overparameterization (Bates, Kliegl, et al. 2015). This yielded the following equation for the model:

$$\begin{aligned}
 & \textit{Systematicity} \sim \textit{functional adaptivity} * \textit{referent environment} \\
 & + \textit{trial number} + \textit{markedness} \\
 & + (1 + \textit{functional adaptivity} + \textit{trial number} \mid \textit{subject/team/condition}) \\
 & + (1 + \textit{functional adaptivity} + \textit{referent environment} + \textit{markedness} \mid \textit{stimulus}) \\
 & + (1 \mid \textit{coder})
 \end{aligned}$$

To control for possible confounds of pair familiarity (whether the two partners knew each other in advance) and pair gender (male, female, mixed), we included these factors in a follow-up model and assessed whether i) the model improved its likelihood via BIC-based model comparison (Burnham & Anderson 2004), and ii) in that case, whether the previous results held. BIC is an index of generalizability of the results, which calculates the likelihood of the model given the data and penalizes it according to the complexity of the model, thus decreasing the risk of overfitting (Brewer, Butler & Cooksley 2016). It is important to note that BIC-based model comparison does not involve p-values and therefore is not as prone to false positive inflation as a statistical significance-based model comparison. All models were implemented with the lme4 package (Bates, Mächler, et al. 2015) in the R environment (R Core Team 2019).

In order to better assess predicted effects that were non-significant in our model, we calculated the Bayes Factor ((Wagenmakers et al. 2010)) to assess whether the data supported our original hypothesis or whether, to the contrary, the null hypothesis would be a better explanation of our empirical findings. We used the Savage-Dickey method and weakly sceptical priors for the effects (normal distribution centred at 0, with a standard deviation of 1). The full scripts and dataset (Nölle et al. 2018a) can be viewed at <https://osf.io/h5eas/>.

2.2. Results

2.2.1. Inter-coder reliability

Inter-coder reliability was found to be between 94 and 97 percent between the different pairs of coders, corresponding to a Cohen's $\kappa = 0.81-0.88$.

2.2.2. Accuracy and conventionalization

The overall mean accuracy was high at 98 percent in both conditions, indicating that pairs were communicatively efficient already from the start of the experiment. Accuracy started at a mean of 93 percent on the first trial and increased significantly ($p < .001$) to over 99 percent on the last trial. By contrast, the number of gestures produced per turn decreased significantly ($p < .001$) - starting at a mean of 4.3 gestures in the first trial and decreasing to an average of 1.3 gestures by the end of the experiment (after 112 trials) with no difference between conditions ($p = .04$, see Figure 2-4).

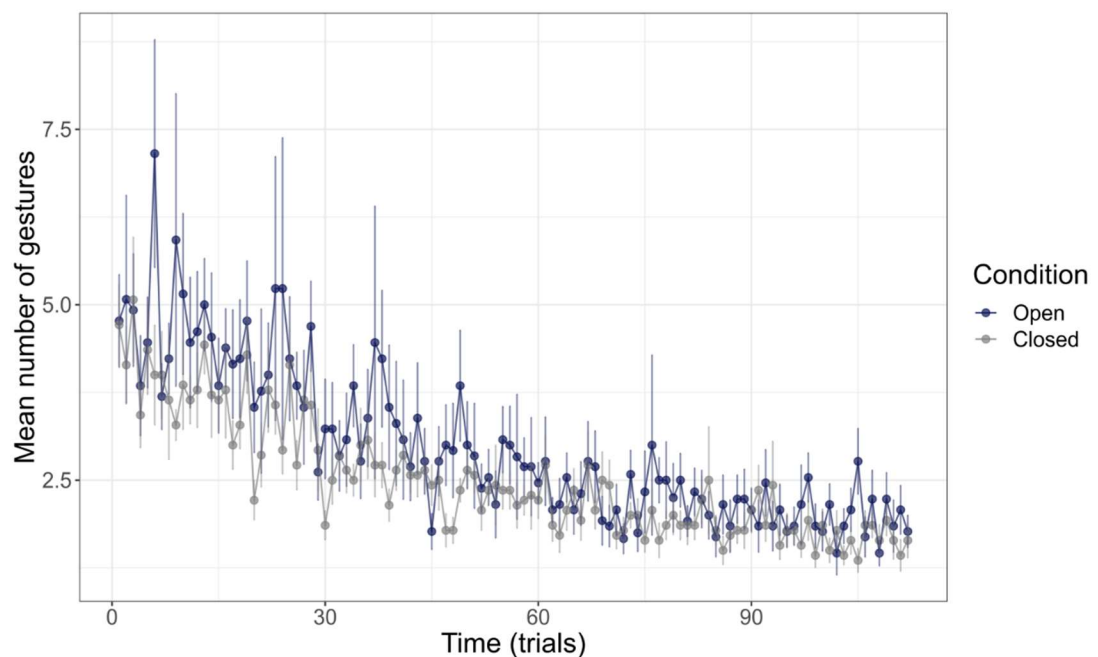


Figure 2-4: The mean number of gestures per trial in the open and closed environment conditions.

2.2.3. Systematicity

Our model revealed a significant main effect of functional adaptivity ($p < .01$), as well as a significant main effect of time (trial number) on systematicity ($p < .0001$). In line with our prediction, traits that were distributed to be highly functionally adaptive over the meaning space were more likely systematized (14.1% probability on any given trial/trait) than traits with low functional adaptivity (2.7%). Contrary to our prediction, the probability of systematic gestures decreased over time (from 10.0% to 0.7% over 112 trials).

No significant effects of referent environment (open vs. closed), markedness or interactions between referent environment and functional adaptivity were found in the data. A Bayes Factor analysis yielded inconclusive evidence for all null results. Adding pair familiarity and gender improved the likelihood of the model but did not lead to any changes in the original results. Table 2-1 presents an overview over the different effects in our model. Figure 2-5 illustrates the main effect of functional adaptivity.

Table 2-1: Study 1 regression model predicting the amount of systematicity used per trial. Note: * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$. β s represent log odds from the logistic regression model. $R^2_{(conditional)} = .78$; $R^2_{(marginal)} = .11$, corresponding to a Pearson's coefficient of .33.

measure	β	SE	Z	BF
functional adaptivity (high=1, low=0)	1.789	0.610	2.93 **	7.99
referent environment (open=1, closed=0)	-0.512	0.638	-0.80	0.70
trial number (z-standardized)	-0.804	0.124	-6.49 ***	> 1000
markedness (marked=1, unmarked=0)	0.770	0.573	-1.34	0.98
functional adaptivity \times referent environment	-0.300	0.813	-0.36	0.36

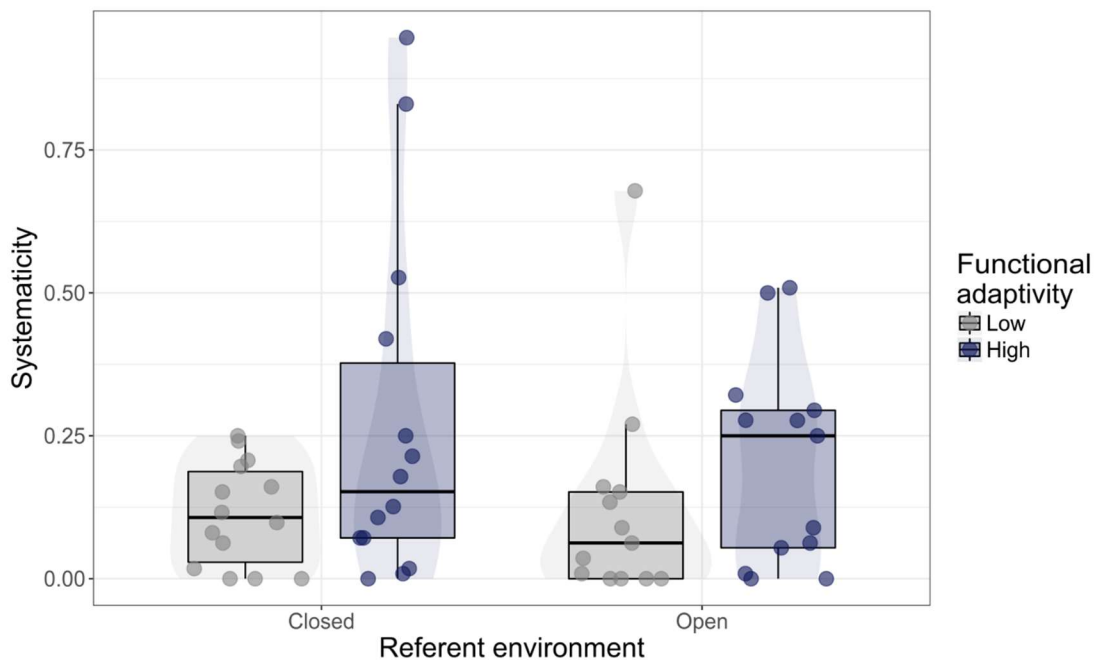


Figure 2-5: Mean proportion of trials with systematic gestures by functional adaptivity and referent environment. Dots represent proportions of systematicity produced by individual pairs.

Note also that the mean is a very imprecise model for this data, as 1) the outcome variable (systematicity) is binary and 2) inter-individual differences are large.

2.3. Discussion

Like in previous studies within experimental semiotics, participants in our experiment readily evolved an efficient and expressive communicative sign system to meet the affordances of the referential task in the absence of a previously established communicative conventions (Galantucci 2009; Galantucci & Garrod 2010; Galantucci & Garrod 2011; Galantucci, Garrod & Roberts 2012). More importantly, in support of hypothesis 1, we found that structurally distributional properties of the environment, in this case, whether referents can be reliably disambiguated with reference to pet or gender, influence whether or not a trait is picked up and systematized in the sign system. In other words, proto-grammatical features of the emergent gestural communication system seem contingent on structural properties of the referential environment that the participants communicate about.

We did not find support for the prediction that “open” environments (with foils changing on each trial) afford more systematic signs to emerge. However, we suspect this result to be related to the particular experimental operationalization: As the target stimuli to be communicated still came from a closed set of 14 recurring items, the affordance for systematicity is likely to have been weak. In fact, debriefing revealed that most participants were aware that certain foils were never designated as targets to be communicated, which narrowed down the search space instead of enlarging it.

Lastly, contrary to our predictions, systematicity did not increase, but rather decreased over time during the experiment. Given this initial experimental setup, it thus appears that over repeated interactions communicating a limited set of referents with the same partner, the most efficient strategy is to use highly idiosyncratic, iconic and simplified signs that can be realised in a single gesture.

However, since we identified some limitations in the experimental design that yield the latter observations potentially confounded, we decided to run a second experiment. Here we operationalized the ‘openness’ of the environment in a more genuine way with new target referents being continuously introduced. Besides, Experiment 2 tests hypothesis 3, that contexts of displaced communication motivate systematicity due to increased working memory loads.

3. Experiment 2: Systematicity emerges in displaced interactions and expanding meaning spaces

3.1. Methods

3.1.1. Participants

40 pairs of participants (8 m, 13 f, 19 mixed, mean age 22.5, SD=2.1) who had not been part of experiment 1 were recruited among students at Aarhus University and took part in the experiment in return for monetary compensation. Pair members did not know each other in advance. All were native speakers of Danish.

3.1.2. Materials, design and procedure

The same materials and procedure were used as in experiment 1. Eight additional stimuli items (representing two additional job categories: clowns and cave people) were added to increase the set of possible referents. The study employed a two-by-two factorial design with referent environment (open vs. closed) as the first factor and communicative context (displaced/co-present) as the second factor. In contrast to experiment 1, the 'open referent environment' condition in experiment 2 sampled both targets and foils from the full set with no repetition of targets within participants. In each trial, the set consisted of 16 randomly sampled stimuli items. In the closed condition the set remained the same for all trials, while in the open condition it kept changing with the constraint that each set had to contain characters from at least six different professions. This was done to avoid highly imbalanced sets, which could confound communicative behaviour. In the closed referent environment condition, participants played two rounds, during which each pair member communicated each stimulus once. In the open referent environment condition, each stimulus from the full set was communicated once by each pair member, yielding a total of 64 trials in both conditions.

The displacement condition was intended to simulate situations where participants communicate about referents not present in the moment of communication. A trial in the displaced context condition thus proceeded as follows (Figure 2-6): The gesturer (but not the guesser) had unlimited time to examine the full set and target stimulus (framed in red). When she was ready to proceed, she pressed a button which made the stimuli disappear from the screen while initiating a 3s countdown, after which she had again unlimited time to produce gestures (in absence of the stimuli). It was then the guesser's turn to indicate when he was ready to make a choice via a button press, again followed by a 3s countdown and finally the full set being displayed to the guesser who then had to make a selection without receiving or producing

any further gestures. In the simultaneous communicative context, the procedure was the same as in experiment 1, with the set being visible to both partners at all times (see Figure 2-6).

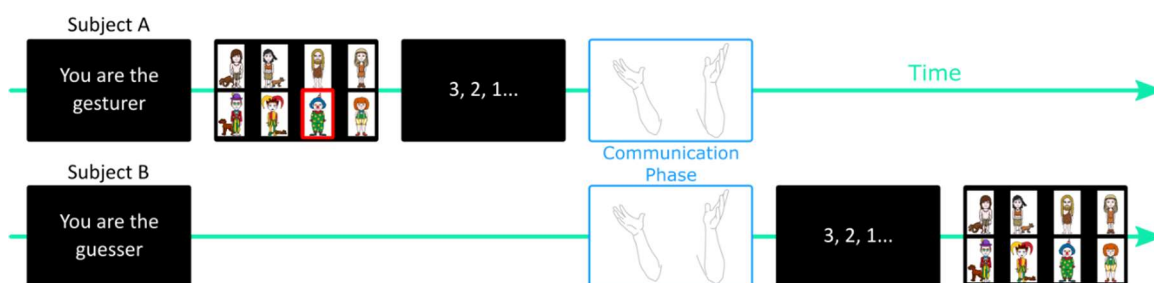


Figure 2-6: Schematic representation of the procedure for a trial in the ‘displaced communicative context’ condition.

3.1.3 Analysis

The same coding scheme and method for quantitative analysis were used as in experiment 1. An independent coder, blind to the hypothesis, coded 75 percent of the data and JN coded 35 percent (i.e. an overlap of 10 percent). To test hypothesis 2 - 3 we built a factorial logistic regression with systematicity as outcome and displacement and referent environment (including interaction term) and time as predictor variables. As traits were randomly distributed in experiment 1, the kind of trait systematised (systematic job/gender/pet gestures) was included as a random effect in the model. Again, as in experiment 1, random slopes for interactions were not included for simplicity. Only team gender was added as an additional control variable in the follow-up model, as none of the participants in experiment 2 knew each other in advance.

$$\begin{aligned}
 &Systematicity \sim displacement * environment \\
 &+ (displacement + environment) * trial\ number \\
 &+ (1 + trial\ number | team/guesser) \\
 &+ (1 + environment + displacement | stimulus) \\
 &+ (1 + environment + displacement | trait) \\
 &+ (1 | coder)
 \end{aligned}$$

The full scripts and dataset (Nölle et al. 2018a) can be viewed at <https://osf.io/h5eas/>.

3.2. Results

2.2.1. Inter-coder reliability

Inter-coder reliability was found to be 97 percent, Cohen’s $\kappa = 0.92$.

2.2.2. Accuracy and conventionalization

Again, as in experiment 1, overall mean accuracy was high, at 95 percent - starting at a mean of 88 percent on the first trial and increasing significantly ($p < .001$) to over 99 percent on the last trial. Accuracy

ranged from 93 to 97 percent in the four conditions, with slightly (but not significantly, $p > .1$) different mean accuracy levels in the open (94%), compared to the closed (96%) environment conditions. The number of gestures produced per turn decreased significantly ($p < .001$) over time - starting at a mean of 5.2 gestures in trial 1 and decreasing to approximately 3.2 gestures. However, there were also condition specific effects with generally more gestures in the open conditions yielding a significant interaction between environment and time ($p < .0001$, see Figure 2-7 and supplementary material).

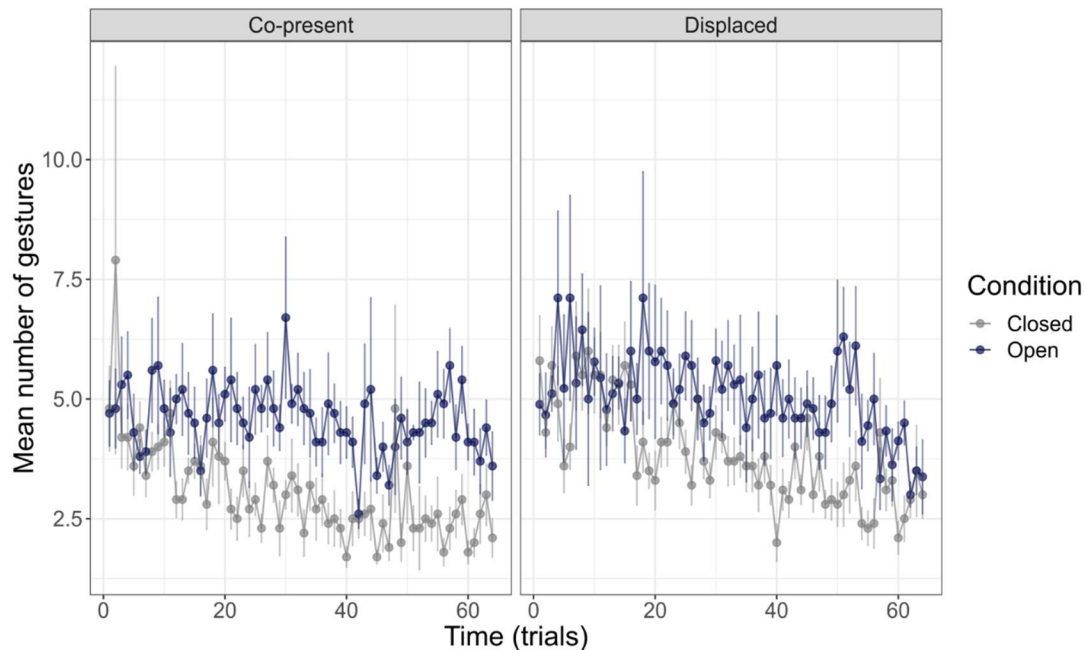


Figure 2-7: The mean number of gestures over time as a function of the two experimental manipulations.

2.2.3. Systematicity

Our model revealed a significant interaction effect between referent environment and trial number ($p < .01$) as well as a significant main effect of displacement ($p < .05$). All other effects were non-significant. A Bayes Factor analysis indicated evidence in favour of the null hypothesis for the main effect of trial and the interaction between displacement and trial, and inconclusive evidence in all other cases (including the main effect of displacement). Adding team gender as a covariate did not improve the likelihood of the model given our data. Table 2-2 presents an overview over the different effects in our predicted model. Figure 2-8 illustrates the effects of displacement and referent environment.

Table 2-2: Table 2. Study 2 regression model predicting the amount of systematicity used per trial. Note: * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$. β s represent log odds from the logistic regression model. $R^2_{(conditional)} = .35$; $R^2_{(marginal)} = .03$, corresponding to a Pearson's coefficient of .17.

measure	β	SE	Z	BF
displacement (displaced=1, co-present=0)	0.886	0.451	1.97	1.16
meaning space (open=1, closed=0)	0.566	0.478	1.18	0.67
trial number (z-standardized)	-0.074	0.083	-0.90	0.10
displacement \times meaning space	-0.488	0.463	-1.05	0.48
displacement \times trial number	-0.031	0.094	-0.33	0.11
meaning space \times trial number	0.254	0.094	2.71 **	2.06

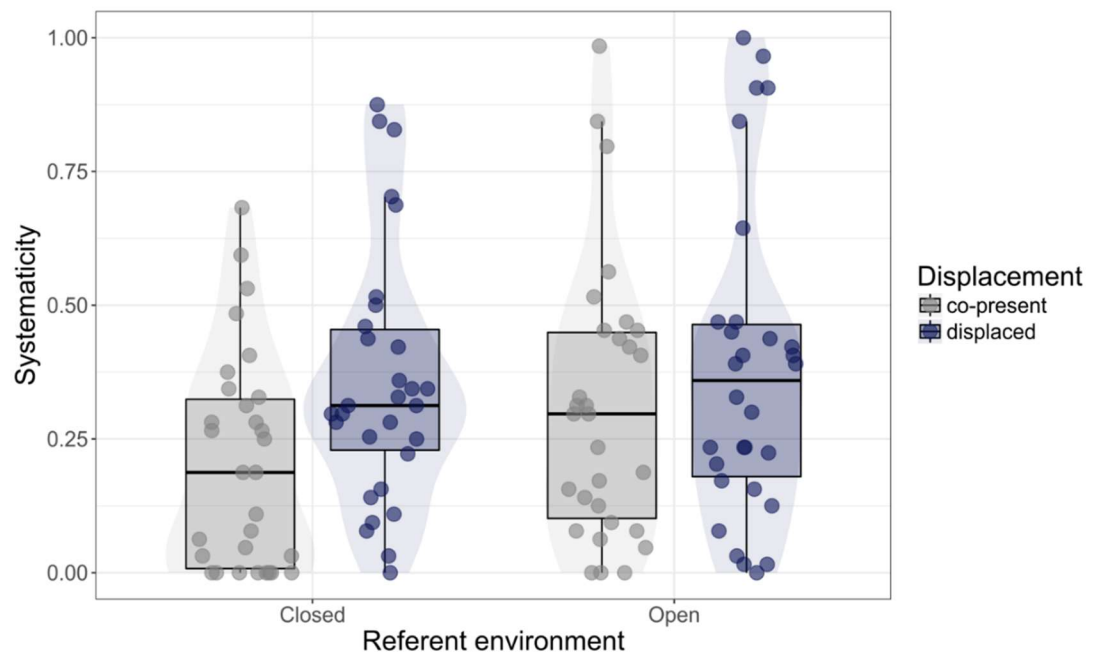


Figure 2-8: Mean proportion of trials with systematic gestures by environment and displacement condition. Dots represent proportions of systematicity produced by individual pairs. Note also that the mean is a very imprecise model for this data, as 1) the outcome variable (systematicity) is binary and 2) inter-individual differences are large.

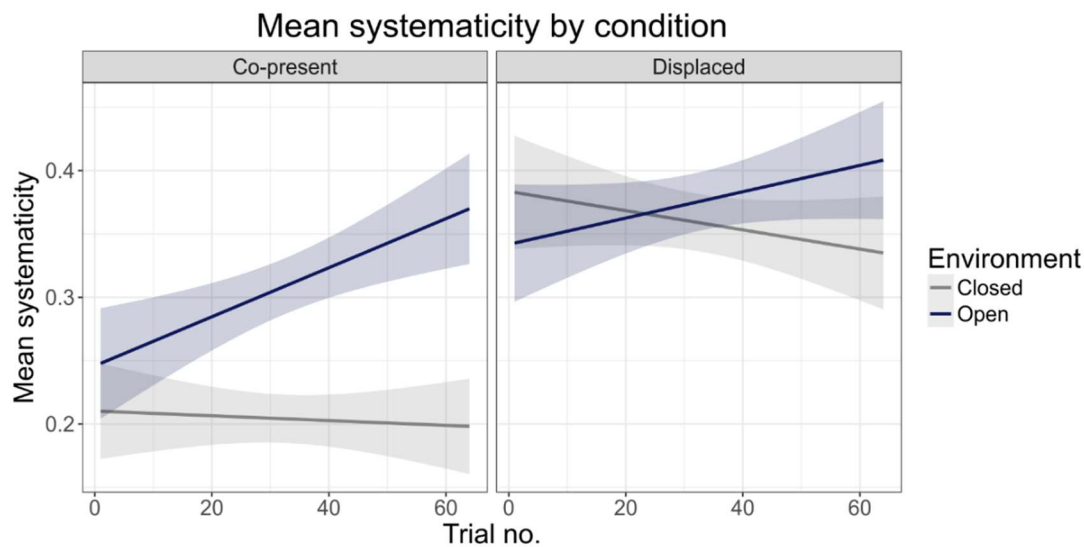


Figure 2-9: Mean proportion of systematic gestures by trial number. Shaded areas represent between-subject standard errors, based on individual pair mean proportions. Note also that the mean is a very imprecise model for this data, as 1) the outcome variable (systematicity) is binary and 2) inter-individual differences are large.

3.3. Discussion

Again, as in experiment 1, participants readily and successfully established an expressive sign system suited to solve the task accurately, and signs became compressed (i.e. simpler) and conventionalized over time. However, the compression was more pronounced in conditions with less systematicity (the closed, co-present condition). With an idiosyncratic strategy, the communication of a referent can potentially be reduced to a single gesture (depicting a single, maximally discriminative feature), while a fully systematic strategy requires the combination of three gestures, one for each of the general features (job, gender and pet). In accordance with hypothesis 2, we observe stronger propensities to systematize signs in the open-environment condition than in the closed referent environment condition. This effect is modulated by time, as participants experience new referents being continuously added to the meaning space. Similar experimental observations have previously been made (Carr et al. 2017; Selten & Warglien 2007) supporting the prediction that systematicity evolves adaptively in response to affordances for productivity and flexibility related to unstable, changing or expanding environments.

We also found significant effects of displacement. That is, participants showed higher propensities to use systematic signs when communicating about referents not present in the moment of communication, thus supporting hypothesis 3. We speculate that such displaced contexts of communication afford participants to abstract and compress information from the immediate visual input into retainable categories more easily sustained in working memory. Unlike open referent environments, the effect of displacement

is a main effect and displacement did not interact with time (trial number). In other words, displacement does not lead to a different trajectory of sign evolution over time but shifts the overall probability of systematic signs being used from the beginning of the experiment. This suggests that - similarly to learnability in cultural transmission studies (Tamariz 2017) - displacement poses compressibility demands on emergent sign systems (Christiansen & Chater 2016). Against our predictions, we did not find significant interaction effects between displacement and open environments. However, importantly, the evidence for both main effect and the interactions remains inconclusive at best as indicated by Bayes Factor.

4. General discussion

4.1. Systematicity as shaped by the environment

Here we have presented experimental investigations directly testing the hypothesis that systematicity evolves adaptively in response to external environmental conditions. In two experiments, we investigated the conditions under which new sign systems evolve to show elements of systematicity. A number of studies have suggested systematicity to be contingent mainly on cognitive constraints related to learning in contexts of inter-generational, “vertical” transmission (Kirby, Cornish & Smith 2008; Kirby et al. 2015; Motamedi et al. 2016; Winters, Kirby & Smith 2015; however, see also Selten & Warglien 2007; Theisen, Oberlander & Kirby 2010). In the present study, we show that systematicity can also emerge in contexts of interactive “horizontal” communication in response to particular structural affordances of the environment and the nature of the communicative context. That is, rather than delegating motivation for systematicity alone to internal cognitive biases and constraints, we suggest that systematicity is the adaptive solution to particular ‘external’ environmental and communicative conditions.

Our findings resonate with a few existing studies on the emergence of systematicity (Motamedi et al. 2016; Theisen et al. 2010). Importantly, in both studies participants’ task was to communicate linguistic concepts belonging to broader categories, such as “hair dresser”, “hair salon” or “fire station”, “fire truck”. Not unlike our studies, systematicity was observed to emerge as reuse of sign elements between items belonging to the same categories. Concerns could be raised, however, that the explicit repetition of forms in the linguistic word stimulus (e.g. “fire”) across several items directly prime systematic repetitions in participants’ communicative responses. Similar concerns can be raised with stimuli used in for instance Kirby et al (2008) and Selten and Warglien (2007) where the highly schematic, combinatorial nature of visual referents constitute a strong prime for systematic responses, while no other solutions are saliently available. In the experiments presented here, categories (JOB, GENDER and PET OWNERSHIP) were more implicit and had to be “discovered” by the participants as a resource for their communicative efforts. In other words,

participants had to simultaneously derive perceptual categories (potentially cutting across multiple individual referents) and invent signs to refer to them. Furthermore, our visual stimulus referents supplied participants with competing affordances for idiosyncratic and systematic strategies, which could be argued to yield more ecological validity.

While the emergence of systematicity has been subject of some experimental investigations, less attention has been directed at what aspects of signs are selected for systematicity and what motivates this. In experiment 1, we manipulated distributional properties of referents (whether the referent characters had a pet or was of a specific gender) in order to investigate if the structure of the referent environment affects which semiotic properties are systematized. We found that evenly distributed features with greater discriminative value are more likely to get picked up and systematized in an emerging sign system. This is in line with Christensen and colleagues' (2016) finding that word order in emerging sign systems is motivated by referent event types through principles of "structural iconicity". Experiment 1, however, goes beyond structural iconicity, as it shows that both the experienced structure of the environment, and the interplay between environment and cognitive efficiency can influence the organization of a communication system: Features with higher informational value (i.e. dividing the search space in a maximally efficient way) are more likely to be systematized, and thus contribute to the communicative efficiency of the sign system.

In experiment 2, we found that the relative stability of the environment influences the development of systematicity over time: Whereas systematicity is unaffected - or even decreases - in stable environments with narrow meaning spaces, open environments with continuously expanding meaning spaces afford for systematic communicative solutions. Combinatorial or compositional signals are more productive and thus are easier to generalize to new referents in the environment than iconic and idiosyncratic form-meaning mappings. In other words, systematicity makes it possible to communicate new referents by re-using and recombining previously established signals thus maintaining the same cognitive economy as in closed meaning spaces. Related observations have been made in an experiment designed to model the costs and benefits of communication (Selten & Warglien 2007): When introduced to a larger set of novel referents, pairs of participants who were able to establish a compositional grammar in previous rounds achieved significantly higher gains than those relying on arbitrary mappings or non-compositional grammars (Raviv, Meyer & Lev-Ari 2019). Importantly, while in Selten and Warglien (2007) and Raviv et al (2019) no competing strategies were practically available, the present study shows that the productivity advantage remains even when idiosyncratic iconic mappings are possible and communicatively effective (i.e., leading to accurate target identification).

Furthermore, we found systematicity to be afforded in contexts of communicative displacement. Although evidence is inconclusive, participants seemed slightly more inclined to use systematic strategies when they communicated about referents not immediately present at the time of communication. Displacement is widespread in everyday linguistic communication and has been proposed as one of the fundamental design features of language (Hockett 1960). Bickerton (2009) even speculates that the ability to communicate about absent or even non-existing referents constitutes one of the most profound achievements in the evolution of language. Interestingly, however, it has not previously been manipulated in experimental studies of the evolution of communication systems (however, see Tamariz & Kirby 2015 for a non-communicative manipulation of displacement). We hypothesize displacement-related systematicity effects to be contingent on human memory bottlenecks: as representations have to be sustained in memory (in this case only for 3 seconds), they are subject to processes of informational compression and biased reconstruction (Bartlett 1932; Kirby et al. 2015; Tamariz & Kirby 2015). While stochastic idiosyncrasies can be more economic to produce (cf. the lower number of gestures in closed and co-present conditions), they are costlier to encode in memory, since encoding processes are facilitated by structure and rule-bound order (Christiansen & Chater 2016). Compression of elements reduces complexity and enables chunking (Chekaf, Cowan & Mathy 2016; Mathy & Feldman 2012), which in turn allows encoding elements for subsequent retrieval from long-term working memory (Ericsson & Kintsch 1995). Consequently, it is easier to sustain communicative target referents to be communicated as systematic compositional representations than as idiosyncratic tokens. While the potential impact of displacement can be argued to be a 'cognitive' effect (since related to memory), it is brought about by changing the pragmatic context from a situation of communicating about co-present referents to one of communicating about referents absent in the moment of communication.

Could the observed effects be driven by other and perhaps simpler factors? In Theisen et al. (2010), systematic responses were observed to often occur when referent stimuli from the same category would appear in immediate succession (e.g. 'school bus' and 'class room'). Curiously, in their study, 21 out of 26 of the initial trials would contain such repetitions of items from the same categories. In these cases, the systematic reuse of items seems motivated by a simple principle of precedence (Garrod & Pickering 2004; Pickering & Garrod 2004). In our study, the referent stimuli were randomized in ways making these kinds of repetitions quite infrequent. Yet, in order to control for a simple precedence explanation of emergent systematicity behaviour in our experiment, we tested if consecutive occurrences of same JOB, GENDER or PET categories would yield systematic solutions more likely of the corresponding traits than cases without such immediate repetition. In experiment 1, none of the three features came out significant ($p > .2$). That is,

participants were not more likely to systematize a trait (e.g. a JOB category) when repeated in two consecutive trials. In experiment 2, JOB and PET again came out insignificant, while GENDER was marginally significant ($p=.04$). However, this seemed driven by later trials. Furthermore, since the number of cases of consecutive trials sharing the same traits were balanced between conditions, a simple precedence factor cannot account for the condition related differences in systematicity reported here.

4.2 Time

We hypothesized that systematicity would interact with time, with systematicity increasing over trials. Across experiments, such effects are only observed in the open referent environment condition. In other conditions, systematicity was either stable (displacement condition) or even decreased over time (experiment 1). Similar observations are made by Theisen and colleagues (2010), who also report main effects of systematicity in graphical dyadic interaction, with no interactions with time. In the closed referent environment conditions, we rather observe participants through repeated interactions condensing initial elaborative sign ensembles to map minimally discriminative details of the target stimuli, thereby often ending up with a single distinct, non-compositional sign for each referent. This corresponds to general observations in experimental semiotics and shows how, as common ground is build up through repeated “horizontal” interactions among interlocutors, signs become less articulated and simplified (Clark & Wilkes-Gibbs 1986; Fay et al. 2010; Garrod et al. 2007), and thus more “cost-efficient” (i.e. demanding less time and energy to produce), consistent with the proposal of Selten and Warglien (2007). However, along with such processes of compression, signs also tend to gradually shift from iconic towards becoming more symbolic, which makes them less semantically transparent and accessible to newcomers. In other words, such strategies come at the cost of reduced learnability and transparency (Fay & Ellison 2013). Vertical transmission (over generations of speakers/signers) on the other hand tends to promote learnability and therefore usually leads to increase in systematicity and regularity (Carr et al. 2017; Kirby, Cornish & Smith 2008; Smith & Wonnacott 2010; Verhoef, Kirby & De Boer 2014). In another recent study, Motamedi and colleagues (2016) found that while (horizontal) interaction leads to communicative efficiency, (vertical) transmission over multiple generations poses compressibility demands which leads to systematicity. In their view, only the combination of interaction and transmission leads to the emergence of both communicatively efficient and structured sign systems that resemble real language (Kirby et al. 2015; Regier, Kemp & Kay 2015). Findings from experiment 2, however, indicate that in contexts of open referent environments, systematicity evolves also gradually in horizontal interaction even if it comes with increased production cost as systematic gestures are more elaborate (e.g., comprising combinations of gestures for GENDER, JOB and PET).

4.3 An ecological view of language evolution

Previous literature has emphasized “inner” cognitive constraints - in particular learner biases amplified through transmission - in explanations of the emergence of systematicity in language (Christiansen & Chater 2008; Griffiths & Kalish 2007; Kirby, Dowman & Griffiths 2007; Kirby, Cornish & Smith 2008). In the present experimental approach, we show that systematic structure emerges in response to broader environmental and contextual affordances. While cognitive processes and bottlenecks are certainly involved, the biases giving more specific, directional shape to the evolution of communication systems need not be delegated to latent inner (and thus innate) factors. Outside laboratories, environments are generally open, large and unstable (see also Carr et al. 2017), with natural, cultural and social changes affording flexibility on all levels of sign use (discourse, lexicogrammar, phonetics). Furthermore, contexts of, for instance, collective foraging could possibly have presented strong affordances for displaced communication that again could have shaped our communicative practices towards becoming more systematic (Bickerton 2009). The morphological features selected for systematicity seem to be contingent on the semiotic potentials of the surrounding environment, that is, the structures stabilizing in evolving communication systems are not independent of the structural organization of the environment that we communicate about (as posited, e.g., by proponents of generative grammar, Chomsky 1986; Hauser, Chomsky & Fitch 2002; Pinker 1994). For instance, if gender provides itself as a reliable trait that can be exploited as semiotic resource for disambiguating referents, we are likely to see this trait getting grammaticalized.

A broader implication of this ecological contextualization of language evolution is that rather than merely presupposing language to be the product of a biological restructuring of the brain (Berwick & Chomsky 2016), or an adaptation to fit the architecture of the biological brain (Christiansen & Chater 2008), it reflects adaptive cognitive strategies responding to external affordances and pressures manifest in the social or physical environment (Everett et al. 2015; Lupyan & Dale 2010, 2016).

5. Conclusions

In two experiments, we tested the hypothesis that systematicity in language evolves adaptively in response to external environmental conditions. Other experimental studies have suggested individual learning mechanisms and cognitive biases related to inter-generational transmission as the major driving force behind the emergence of systematic structure (Kirby et al. 2015; Motamedi et al. 2016; Tamariz & Kirby 2016). The experiments presented above extend these findings showing that systematicity emerges also in contexts of horizontal interaction in response to aspects of the communicative situation itself, even in the presence of competing iconic solutions observed to be effective in several previous experimental studies (Fay, Arbib & Garrod 2013; Garrod et al. 2007; Perlman, Dale & Lupyan 2015).

We have shown that the selection of traits to be systematized depends on their functional adaptivity in the given environment, leading to co-emergence of systematic categories and signals to encode them. In addition, environmental factors also affect the propagation of systematicity in emerging communication systems to the extent that they pose informational bottlenecks. Systematicity is thus more adaptive in an unstable, changing and open environment where analytic forms give flexibility and productivity to communicate about novel referents.

Furthermore, displacement of referents from the communicative context could exert working memory pressures making systematic, compressible structure more favourable. Although displacement is a fundamental property of language (Hockett 1960), it has been largely neglected in laboratory studies investigating cultural language evolution that have rather emphasized learning. Manipulation of displacement in artificial language learning tasks thus presents itself as a promising avenue for further research in the field.

Acknowledgements

We would like to thank the Cognition and Behavior Lab at Aarhus BSS for facilities and equipment as well as Sergio Gonzales De La Higuera Rojo and Caroline Kildahl for coding of the video material. This work was funded by a Seed Grant of the Interactive Minds Centre as well as funding from the School of Communication and Culture, Aarhus University.

Appendix A. Supplementary material

All datasets, analysis scripts and stimuli material can be accessed at <https://osf.io/h5eas/>.

2.3 Summary of Paper 2

Paper 2 presented two silent gesture experiments showing that the emergence of systematic structure in a novel communication system can be affected by external variables that relate to the referential environment and communicative context. In these studies, systematicity encompasses several elements: Firstly, to become systematic, signs had to be recurrent in that the same form was used to refer to another referent. In this way, a gesture for *female* describing a FEMALE SOLDIER could be extended to refer to, e.g., a FEMALE CHEF. Or a gesture expressing *male* could be combined with a gesture that could express *chef* in the latter example. In this way, systematic signs appeal to semantic relations between their referents and involve an element of abstraction in that they allow to create and subsequently mark abstract categories such as GENDER or OCCUPATION. The next important element is conventionalisation: Once, both members of the dyad agree on a sign to mark a category and start using it to refer to several referents, it becomes productive

and can be combined in a compositional way with other signs, similar to how case or gender markers function in natural languages.

In the present study, systematicity furthermore interacted with iconicity. Since referents (cartoon images) and the communication medium (manual gesture) were both visual, there was a strong affordance for grounding iconic signs based on idiosyncratic traits. Iconic signs are thus motivated, as they are based on a resemblance between meaning and form. During the task, an idiosyncratic iconic sign could, for instance, pick up on a single unique trait present in a target image (such as a moustache or crown). By contrast, a systematic sign can also remain iconic (such as a petting gesture to refer to images that depict a cartoon figure with a pet) but does not disambiguate the referent based on a unique feature. In this way, such a petting gesture can be used to refer to all cartoon characters that are depicted with a pet, regardless of whether it is a cat or a dog. In these silent gesture experiments, we are thus dealing with motivated systematic signs. These signs remain iconic in the sense that they resemble their meaning and are to some extent transparent, but they can be called systematic insofar as they represent a more abstract category and can be combined with other signs. In natural languages, motivated systematicity arises from sets of iconic words sharing systematic relations, such as ideophones that often share structures like reduplication, as seen in Japanese (Dingemanse et al. 2015), or phonaesthemes like *glisten*, *glare*, *glimmer*, *glaze*, *glow*, etc. (Bergen 2004). We also find *non*-motivated systematicity in natural languages. Here, similar (non-iconic) word forms are mapped onto similar meanings, e.g., in the case of *wh*-question words (Slonimska & Roberts 2017) or the largely arbitrary phonological structure of verbs vs. nouns (Monaghan, Christiansen & Chater 2007). Finally, a large portion of the lexicon of natural languages consists of purely arbitrary signs that are both non-motivated and non-systematic. In other words, iconicity and systematicity interact in different ways (summarized in Table 2-3 below) and should not simply be subsumed under ‘non-arbitrary’ or ‘motivated’ to distinguish them from arbitrary symbols. Rather, they might be better conceived as orthogonal features of sign systems related to different adaptive pressures (see also Dingemanse et al. 2015; Nielsen 2016).

Table 2-3: Possible types of signs.

	Iconic signs	Arbitrary signs
Idiosyncratic signs	Motivated non-systematic	Non-motivated non-systematic
Systematic signs	Motivated systematic	Non-motivated systematic

The present study demonstrated that the process of picking out recurrent traits from the environment to ground systematic signs is contingent on pressures from the referential environment itself: Traits that were

more salient and functionally adaptive to disambiguate targets were more likely to become expressed as systematic categories. Additionally, the mode of communication (communicating about present or absent entities) and openness of the environment affected the overall degree of systematicity and the stability of systematic signs, respectively. In this way, the study showed both that systematicity can arise in competition with iconicity (or more specifically, motivated non-systematic signs, see Table 2-3) and that the emergence of structure in a communication system is sensitive to the environment and communicative context in which it used.

Taken together, Paper 2 suggests that communication systems evolving in different referential environments could end up with different systematic categories reflecting their functional adaptivity to their communicative or referential environment or 'niche'. To understand better whether such environmental factors also play a role in shaping the diversity of real-world languages, the following chapters will look at spatial language, which provides a good test case for linguistic adaptation due to its complex relationships with culture, cognition and the physical environment.

3 Spatial language: Linguistic adaptation at the intersection of language, culture and cognition

3.1 Introduction: Spatial language and cognition

The domain of space forms a fundamental part of all languages, given that human activity across all cultures takes place in three-dimensional space (Levinson 2003). In consequence, spatial referencing systems have evolved and adapted exquisitely to enable us to talk about orientation, direction as well as distant or topological relations between objects, and spatial meanings often provide basic source structures that more abstract meaning dimensions can be mapped onto (e.g., Lakoff & Johnson 2003).

Notably, while all languages (to my knowledge) provide their speakers with the ability to express spatial relations between entities, spatial language is highly diverse: There are striking differences in how objects are divided intrinsically into functional segments (e.g., “fixed armature” vs. “object-centred” body-part terms; see Levinson 2003: 78), how topological relations (like ON or AT) divide semantic space (Bowerman & Choi 2001), how aspects of motions like PATH, MANNER, GOAL or DEIXIS are lexically encoded (Talmy 1985; Choi & Bowerman 1991), or which demonstratives are used to refer to objects in different distances from the speaker or listener (Diessel 2014; Levinson et al. 2018).

In this thesis, I focus on a single aspect of spatial language: How languages express distant relations, that is the relation of a FIGURE object to a GROUND object based on an anchored coordinate system, a so-called Frame of Reference or FoR (Levinson 1996; Levinson 2003). Frames of Reference have likewise been described as highly variable: While speakers of Western, Indo European languages like English or German usually rely on egocentric (relative) or object-centred (intrinsic) coordinate systems to describe spatial relations between objects, many other languages predominantly employ an absolute or allocentric system based on salient properties of the environment like inclinations, prominent landmarks or cardinal directions (see, e.g., Levinson & Wilkins 2006 for a sample of languages).

There have been many discussions about these coordinate systems, their relation to general cognition, to sociocultural variables, and more generally their origins. I will first give an overview about FoR and the different ways in which they have been classified in the literature. I then discuss how they vary in relation to non-linguistic variables such as cognition and demographic factors as well as with respect to the physical environment. Finally, this chapter includes a paper arguing that if we want to understand

the complex causal relationships and mechanisms shaping spatial language, we have to combine more traditional descriptive approaches that map out the diversity of spatial language with rigorous experimental and computational methods that allow modelling the mechanisms that have shaped and continue to shape preferences in spatial referencing strategies.

3.1.1 Frames of Reference

As noted above, languages utilise different coordinate systems to express the relations between objects that are not in direct contact with each other (see Figure 1), so-called Frames of Reference (FoR). These coordinate systems reflect the basic FIGURE-GROUND distinction guiding human visual perception that is known from Gestalt psychology (Wertheimer 1923; Wagemans et al. 2012) and roughly corresponds to the SUBJECT-OBJECT distinction present in languages (Levinson 2003). Other terms that are commonly used in the literature include REFERENCE (or LOCATUM/THEME) and RELATUM (Miller & Johnson-Laird 1976) or TRAJECTOR and LANDMARK (Langacker 1987).

In addition to FIGURE (F) and GROUND (G), FoR, according to various definitions, also encode information about the ORIGIN of the coordinate system (X), the VIEWPOINT (V) and the ANGLE/DIRECTION between F, G and V (or another anchor). In English, for instance, three distinct ways can be used to describe the position of F (e.g., a *ball*) in relation to G (e.g., a *car*), which are depicted in Figure 3-1 below: Why is it that some languages rely on properties of the environment to express spatial relations and others do not?

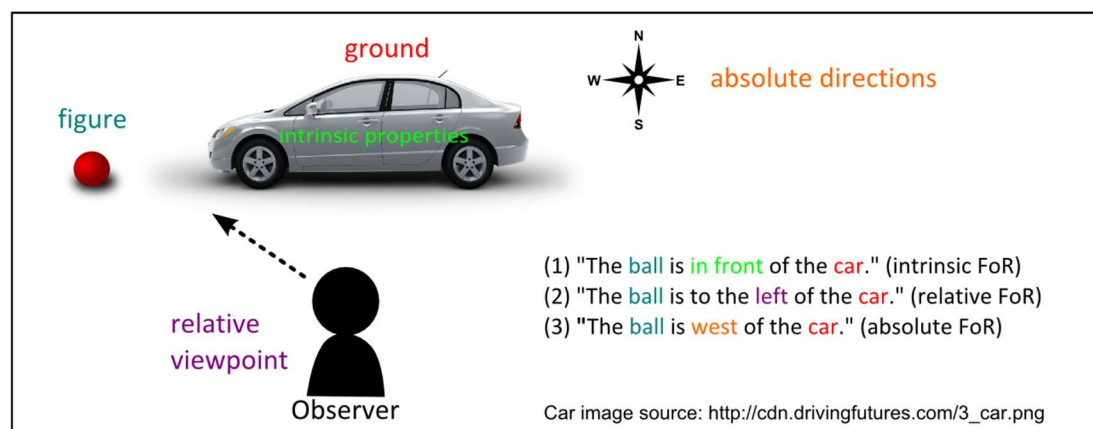


Figure 3-1: Example of a spatial scene described according to three distinct spatial reference systems in English (adapted from Levinson 2003: 25).

In example (1), intrinsic properties of G (the functional *front*) are used to determine F's location. This FoR can be therefore be called 'intrinsic'. In (2), a third coordinate, V, is used to locate F in relation to G. Note, that V is relative, as it can either be the perspective of the sender or receiver, which is why this FoR is called 'relative'. Lastly, the 'absolute' FoR, consists of strategies, where F and G are localized in respect to an

external anchor, which can be a landmark or an absolute direction, like *west* in (3). This distinction was formulated by Levinson (1996, 2003) and focuses on the different numbers of elements involved in the scene and properties of the array under rotation, rather than the origin of the coordinate system.

While Levinson's three-fold FoR classification is probably the most widespread in linguistics, others have proposed alternative numbers of frames: For example, Jackendoff (1996) distinguishes 8 different frames, while many cognitive psychologists (e.g., Klatzky 1998; Galati et al. 2000; Li & Gleitman 2002) rely on a simpler binary distinction between 'egocentric' and 'allocentric', which has also frequently been used to characterise other cognitive (and not necessarily spatial) representations that can be grounded in relation to a self or others, such as in mentalising or perspective-taking, which form part of social cognition (Frith & de Vignemont 2005). Finally, some FoR classifications highlight different aspects of the respective coordinate systems such as the origin/anchor emphasized in the distinction between object-centered (intrinsic), viewer-centered (deictic), an environment-centered (extrinsic), which roughly corresponds to Levinson's intrinsic, relative and absolute FoR (Carlson-Radvansky & Irwin 1993; for further alternative classifications cf. Danziger 2010; Diessel 2014).

Over the years, there have been many disagreements over the correct classifications of these frames (see, e.g., Palmer 2002; Li & Abarbanell 2018), and it can often be hard to compare findings across authors using different categories. In order to unify these multiple accounts, Bohnemeyer and colleagues (2015) have recently proposed a new FoR classification that distinguishes six distinct frames and solves the issue that, for example the popular ternary relative-intrinsic-absolute classification and binary egocentric-allocentric classifications cannot be easily aligned (see Table 3-1). They point out that the psychological distinction focuses on the anchor that defines the axes of the coordinate systems (the observer's viewpoint for the egocentric FoR vs an environmental entity or GROUND object for the allocentric FoR, which could be called geocentric and intrinsic, respectively). By contrast, linguistic typological studies have distinguished between 'intrinsic' and 'extrinsic' frames based on the origin of the coordinate system (Danziger 2010), both of which Levinson subsumes under 'intrinsic'. Bohnemeyer et al. therefore explicitly distinguish between the 'allocentric intrinsic' frame where the anchor lies outside the observer and instead in the GROUND (e.g., when referring to a FIGURE *in front of* or *behind* a GROUND object, such as *the ball in front of the car* in Figure 3-1), and the 'egocentric intrinsic' frame (which Danziger 2010 calls 'direct'), where the GROUND that anchors the coordinate system lies inside an observer (e.g., the observer depicted in Figure 3-1 seeing the ball *to the left* of the car, or an imaginary interlocutor on the far side of the car that would see the ball as *to the right* of the car).

Table 3-1: Slightly modified reference frame classifications from Bohnermeyer et al. (2015: 175). Sources for classifications: a) e.g., Carlson-Radvansky & Irwin 1993; Wassmann & Dasen 1998; Li & Gleitman 2002; b) e.g., Levinson 1996, 2003; Pederson 1993; Danziger 2010; Bohnermeyer & O'Meara 2012.

a) Classification based on anchor*		b) Classification by anchor and origin of axes	
egocentric		egocentric extrinsic = relative (Levinson 1996) anchor = body of an observer ground ≠ anchor axes projected (translated and/or reflected) <i>The ball is to the right of the chair.</i>	relative
		egocentric intrinsic ('direct' in Danziger 2010) anchor = body of an observer ground = anchor axes extended (without projection or abstraction) <i>The ball is in front of me.</i>	
allocentric	intrinsic	allocentric intrinsic anchor ≠ body of an observer ground = anchor axes extended (without projection or abstraction) <i>The ball is in front of the chair.</i>	
	geocentric	landmark-based ('projected' in Mishra et al. 2003; 'head-anchored' in Bohnermeyer and O'Meara 2012) anchor = environmental entity/feature ground ≠ anchor axes defined as vectors pointing toward/away from landmarks <i>The ball is mountainward of the chair.</i>	intrinsic
		geomorphic ('contextual' in Jackendoff 1996: 17) anchor = environmental entity/feature ground ≠ anchor axes projected <i>The ball is downriver of the chair.</i>	
		absolute (Levinson 1996; 'geographical' in Jackendoff 1996) anchor = environmental entity/feature ground ≠ anchor axes abstracted from geomorphic or landmark-based system <i>The ball is downriver of the chair.</i>	absolute

Note that from this chapter onwards, all FoR terminology in this thesis refers back to this classification shown in Table 3-1. This means that for instance, 'allocentric' is not to be understood synonymously with 'geocentric' or 'absolute'. Rather if I use the term 'allocentric', I intend a collective reading that includes the 'allocentric intrinsic' frame as well as the 'geocentric' frame (including both 'landmark-based' and 'geomorphic' in Bohnermeyer's classification). By contrast, 'geocentric' excludes the 'allocentric intrinsic' frame, and 'absolute' only refers to systems using abstracted axes that behave similarly to the cardinal directions. Likewise, if I use the term 'egocentric', I collectively refer to both the 'relative' frame in

Levinson's sense and the 'egocentric' intrinsic frame in Bohnermeyer's sense, but not the 'intrinsic' frame in Levinson's sense, which would include the 'allocentric intrinsic' frame.

3.2 Origins and diversity of FoR

How are these FoRs distributed across the languages of the world and where do they originate from? According to Levinson, all languages use FoRs that correspond to his ternary classification, although curiously, languages vary in how many FoRs are present and which frames are predominantly used (see Levinson & Wilkins 2006 and Majid et al. 2004 for samples of languages). For example, while English allows spatial descriptions in all three FoRs, the relative and intrinsic FoR are most commonly used in daily language, while the absolute is rarer. Other languages mainly rely on different versions of the geocentric or absolute FoR, and often exclusively use such an environment-based coordinate system, even when referring to relations between objects in small-scale space.

For instance, Guugu Yimithirr speakers (Australia) have been described to solely rely on the cardinal directions (similar to example (3) in Figure 3-1) and Tzeltal speakers (Mexico) primarily use a geomorphic *up(hill)-down(hill)* system, (e.g., translating to something like *the ball is uphill of the car*), relating objects the axes projected from slopes present in the environment (although these may not always be visible). It is important to draw a distinction here, between geocentric systems that are completely invariant to the position of the speaker ('geomorphic' in Table 3-1) and those that rely on the actual presence of a landmark or other environmental cue (slopes, rivers, mountains, shores etc.) that can be directly pointed at ('landmark-based' in Table 3-1). In some communities, environmentally grounded features might become abstracted and functionally fulfil the role of absolute cardinal directions. Earlier descriptions of Tzeltal interpreted the use of *uphill*, *downhill*, and *across* by its speakers in this absolute sense (Levinson 2003; Brown 2006). However, this view has been challenged by Polian and Bohnermeyer (2011) who found that Tzeltal speakers at several field-sites would use the same spatial terms, but would ground them based on different slopes in the nearby environment. Very similar observations have been made by Wassmann and Dasen (1998) for the use of geocentric *seaward* and *islandward* spatial terms in Balinese that vary between field-sites with different coastlines.

Does this mean geocentric systems of this kind emerge due to the presence of salient local landmarks in respect to which they are grounded? Curiously, *up-down* systems that are analogous to Tzeltal (a Mayan language), exist in many phylogenetically unrelated languages spoken in smaller societies and mountainous areas featuring prominent slopes, such as Chintang (Dirksmeyer 2008) and other Kiranti languages spoken around the Himalayas like Belhare (Bickel 1999) or Rai (Ebert 1999), or Swiss dialects in the Alps

(German and Italian-based, see, e.g., Zinsli 1945; and Vicario 2008). In these cases, the cultural importance of verticality (e.g., in cosmology) or its perceptual salience (mediated by the presence of slopes and mountains) appears to be reflected in the local language.

Analogously, depending on the geographic area, other landmarks might be used: Languages spoken in regions with rivers often rely on a river-based system instead, e.g., Jaminjung (Australian/Northern Australia; Hoffmann 2011), Jahai (Mon Khmer/Malay Peninsula; Terrill & Burenhult 2008) or Mian (Ok/Papua New Guinea; Fedden & Boroditsky 2012). Are these cases evidence for environmental adaptation? Based on such observations various scholars have speculated whether geocentric FoR systems are determined by the environment.

Palmer and colleagues have observed that atoll-based languages exhibit a *lagoonward/oceanward* distinction that only exists in languages spoken in this type of environment, such as Marshallese, and Austronesian spoken in Central Micronesia or Dhivehi, an Indo-Aryan language spoken on the Maldives (Palmer 2007; Palmer 2015; Palmer et al. 2017). Notably, Marshallese is also spoken in an urban setting (Springdale/US), where descriptions relying on this geocentric FoR are not found, and the relative FoR is used much more frequently (Palmer et al. 2017; Schlossberg 2018). Intuitively, then, these cases seem to suggest strong evidence for an environmental effect of topography on spatial language.

However, Majid and colleagues (2004: 112) argue against linguistic adaptation, or what they call “simple [ecological] determinism”. They considered 21 languages in relation to three types of variables and their correlation with FoR preference: Environment (urban vs. rural), action (subsistence mode, e.g., hunter-gatherer vs. industrial) and cognitive style (collectivism vs. individualism) and conclude that none of these variables patterns clearly with the choice of a specific FoR, with the exception of a mild relationship between urban societies and the use of the relative FoR.

So what is the right answer? I would argue that there is no simple answer, since the question about spatial language and environment is complicated by four distinct problems:

1. The issue of inferring statistically robust and causal relationships from cross-sectional observational data.
2. The conflation of language and speech-community.
3. The complexity of the relationships between spatial language, cognition and environmental factors.
4. Biases in the literature pertaining to different positions with regards to these relationships.

Problem 1 was already discussed in Section [1.4](#) above and will be discussed in greater detail in [Paper 3](#) below. In the context of this chapter, it should suffice to say that in the past, many conclusions about spatial language and environment were based on studies that tested single causal links, relied on qualitative observations and correlations between variables, and did not control for cultural evolutionary confounds.

Problem 2 relates to many studies in this context that focus on variation in spatial referencing *between* linguistic communities, rather than *within* communities (e.g., Pederson et al. 1998; Levinson & Wilkins 2006; Majid et al. 2004; cf., Palmer et al. 2018; O'Meara & Pérez Báez 2011; Bohnermeyer et al. 2015). For example, in their edited volume, Levinson & Wilkins (2006) present spatial sketch-grammars for twelve languages. Such grammars assume that a 'language' X has a single preference for one or more FoR, which precludes variation within this language. However, in recent years, field-linguists have increased the granularity of their comparative work on FoR: Projects such as the comparative research on atoll languages described above (Palmer et al. 2017; Palmer et al. 2018), or a similar large comparative project on spatial language in Meso-American languages (O'Meara & Pérez Báez 2011; Bohnermeyer et al. 2015) have moved away from treating languages as single datapoints and instead consider individual field-sites, where individual speakers (or dyads) are treated as the smallest unit of analysis, which allows for the consideration of variation within languages. For example, if one were to treat Swiss German or Italian as a single datapoints, most scholars would refer to these as languages that rely mostly on the egocentric and intrinsic FoR, which would ignore the findings from dialects spoken in the Alps that exhibit more geocentric features (see above).

Additionally, moving away from 'languages' as units of analysis also has consequences in relation to **Problem 3**: Spatial FoR have been found to interact with a large variety of non-linguistic factors, and these are not necessarily the same at different field-sites. These factors include among others demographic variables such as age, first language, second language(s), education, gender, subsistence mode (Danziger & Pederson 1998; Pederson et al. 1998; Mishra et al. 2003; Eggleston et al. 2011; Polian & Bohnermeyer 2011; Haun et al. 2011; Palmer 2015; Bohnermeyer et al. 2015; Meakins et al. 2016; Lin 2017; Palmer et al. 2017; Shapero 2017; Donelson 2018; Moore 2018) as well as environmental variables such as urban vs rural dwelling or the geographic presence of topographic landmarks (Pederson et al. 1998; Mishra et al. 2003; Dasen & Mishra 2010; Adamou & Shen 2017; Palmer et al. 2017). Importantly, none of these variables affects FoR preferences across *all* languages or sites. Rather many of these effects seem specifically tied to the unique sociotopographic makeup of a particular field-site or community, which can lead to surprising findings: For instance, Calderón et al. (2019) found that a population of bilinguals in Spanish

and Ngigua (Otomanguean) in rural Mexico prefer the geocentric FoR in both verbal descriptions in both languages as well as accompanying gestures. This appears in line with findings from other contact situations (e.g., Bohnermeyer et al. 2015; Meakins, Jones & Algy 2016). However, Spanish monolinguals in Calderón et al.'s study showed the same preference for geocentric co-speech gestures as their bilingual counterparts. This is surprising, because language contact usually leads to the import of egocentric terms from 'majority' languages like Spanish or English at the expense of geocentric strategies (e.g., Bohnermeyer et al. 2015; Schlossberg 2018), which is in line with the idea of a cultural override of geocentrism via the introduction of egocentrism (Haun et al. 2006). This directly relates to another issue that I have neglected to mention so far: When I talk about 'preferences' in FoR, I am not only referring to linguistic preferences, but also general cognitive and behavioural preferences, such as in the case of the Ngigua co-speech gestures.

This relationship between spatial language and performance on non-linguistic cognitive tasks has, for decades, been at the centre of discussions about diversity in spatial language, since early comparative field-studies by members of the Max Planck Institute in Nijmegen in the 1990's showed that speakers of languages that prefer geocentric descriptions, perform differently on spatial orientation or memory tasks (Levinson 1997; Pederson et al. 1998). This was one of the main findings that led to a revival of the linguistic relativity hypothesis and discussions about the relation of language and thought (most prominently in Levinson's 2003 book). To this day, this Whorfian interpretation remains hotly debated with various groups arguing for or against the idea that language shapes spatial reasoning or that speakers preferring the geocentric FoR experience the world fundamentally differently from speakers that prefer the egocentric FoR (see, e.g., Li & Gleitman 2002; Li et al. 2011; Li & Abarbanell 2018 against; Pederson et al. 1998; Levinson et al. 2002; Haun et al. 2011; and Lucy 2016; as well as Everett 2013a: chap. 6 for overviews).

Problem 4 in understanding the relationship between spatial language and sociotopographic variables is that this debate is so prominent in the literature that it can lead to biased views regarding the role of the environment. For instance, proponents of the Neo-Whorfian view, might align more with a view where the environment does not motivate variation in spatial language (see, e.g., Majid et al. 2004), since environmental influences can be used to construct an argument where the environment determines FoR choice both in language and cognition simultaneously, thus weakening the importance of the relation between the two (e.g., Li & Gleitman 2002).

In my view, this debate should not be a major focus of research into linguistic adaptation. Relationships between linguistic, cognitive, social and environmental variables are extremely complex and hard to

disentangle (see section 1.4). To understand the full picture behind these relationships, we should rather try to uncover the mechanisms that can explain the emergence of this variation in both domains, without making strong assumptions about whether spatial language influences spatial reasoning or vice versa. In the next section, I will present a paper that tries to operationalise such an experimental, evolutionary framework that can be used to tackle causality in spatial language by integrating traditional approaches with experimental methods used in language evolution research.

3.3 Preface to Paper 3

Paper 3, (Nölle & Spranger, accepted) will be included as part of a special issue on *Sociotopography: the interplay of language, culture and environment in representing space* published in *Linguistic Vanguard* and scheduled for release in 2021. The paper was written in collaboration with Michael Spranger (Sony Computer Science Laboratories Inc). I wrote the first draft of the paper and created the figures (except for Figure 4, which was contributed by my co-author). Michael Spranger commented on the final version, gave advice on structure and edited the subsection on computational models that focuses on his work simulating the evolution of spatial language in artificial agents. We also received comments on an earlier version by Alice Gaby and two anonymous reviewers. The present version has been accepted for publication and we are awaiting the final proof at the time of submission of this thesis. The paper presents a theoretical framework that outlines how the adaptation of spatial language can be studied from an evolutionary perspective. The discussion of the role of causality in cultural evolution was also influenced by my involvement with the CHIELD database that we discuss in the paper. During my PhD, I contributed several causal graphs to the database and co-authored the overview article that introduces how CHIELD can be used to formulate and test causal hypotheses (Roberts et al. 2020). The database can be found at <https://chield.excd.org/>. Note that the reference list for Paper 3 can be found at the end of this thesis. Figure numbers have also been changed to be consistent with the rest of this document.

3.4 Paper 3:

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

Jonas Nölle, Michael Spranger

Abstract

There is striking variation 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, e.g., topography, subsistence style, or 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 that lead to selection of one FoR strategy over another are still not fully understood. In this paper, we argue that in order 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 such a framework, causal relationships between linguistic and non-linguistic variables (such as topography and FoR choice) can be isolated and empirically tested in order to shed light on how sociotopographic factors motivate the variation in spatial language we observe cross-linguistically.

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 Lupyán & Dale 2016 for a review). Proposed links between linguistic and environmental variables include the relationships between morphological complexity and population structure (Bentz & Winter 2013; Lupyán & Dale 2010), between climate and sound systems (Everett, Blasi & Roberts 2015), lexical categories (Brown & Lindsey 2004; Regier, Carstensen & Kemp 2016), and even subtle biases resulting from anatomy (Dediu, Janssen & Moisik 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; Majid et al. 2004; Levinson & Wilkins 2006). It has been argued that these differences do not simply constitute cross-linguistic variation, but also affect speakers' general judgements and cognitive strategies on non-linguistic spatial tasks (e.g., Pederson et al. 1998; Levinson 2003; Majid et al. 2004; Haun et al. 2011), although this remains highly controversial (cf. e.g., Newcombe & Huttenlocher 2000; Li & Gleitman 2002; Gallistel 2002; Li et al. 2011; Pinker 2007; Diessel 2014). 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 'natural', but this turned out to be a Eurocentric fallacy resulting from the lack of cross-cultural data (see, e.g., Levinson 2003, chapter 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 & Li 2016). This suggests that geocentric FoR might be 'older', while egocentric systems are a more recent innovation (see also Bohnermeyer et al., this issue).

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 An open question is 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, e.g., 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 socio-topographic variables and their relationships to investigate causal effects and gain a more mechanistic *understanding* of FoR, their origins and variation.

In section 2, we will first review challenges of traditional approaches studying spatial linguistic phenomena, which largely result from the complex relationships between language and socio-topographic

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

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. Finally, we conclude that such a ‘maximum robustness approach’, including fieldwork as well as computational models and 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 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 North-East peninsula in a way that appears to reflect the local topography. However, it has proven difficult to attribute such variation to a causal mechanism. The first larger survey of 21 languages by Majid and colleagues (2004: 112) concluded that there is no discernible pattern suggesting that environment or sociocultural factors such as dwelling or subsistence style would correlate with a specific FoR.

It has since become clear that a major problem of earlier accounts was equating languages with specific FoRs, i.e. 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 being accounted for (Palmer et al. 2018). Factors which have been observed to correlate with specific FoR choices are, e.g., age (Eggleston, Benedicto & Balna 2011; Polian & Bohnemeyer 2011), first language (Pederson et al. 1998; Bohnemeyer et al. 2015; Haun et al. 2011; Donelson 2018), education/literacy (Danziger & Pederson 1998; Mishra, Dasen & Niraula 2003; Lin 2017), gender (Lawton 2001; Ameka & Essegbey 2006; Bohnemeyer 2011; Le Guen 2011), occupation/subsistence (Palmer et al. 2017; Shapero 2017), second language (Palmer 2015; Palmer et al. 2017; Lin 2017; Moore 2018), and various environmental factors such as settlement (Pederson et al. 1998; Mishra, Dasen & Niraula 2003; Adamou & Shen 2017) or local topography (Dasen & 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 & Shen 2017; Calderón, De Pascale & Adamou 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 also Palmer and colleagues’ introduction to this special 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 13 Mesoamerican languages required a large-scale research project involving many collaborators (O’Meara & Pérez Báez 2011). Hence, many descriptions of spatial grammars stem from different research groups using different methods and even theoretical frameworks, e.g., with regard to how FoR are defined and coded (Bohnenmeyer 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 universal claims, should ideally be accounted for in all related studies. *Post hoc*, this is often impossible. For instance, Palmer (2015) changed his ‘Topographic Correspondence Hypothesis’ into a ‘Sociotopographic Model’ after further studies revealed that social variables like subsistence style can modulate the effect of topography (Palmer et al. 2017). 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 (Bohnenmeyer 2016). Cross-sectional datasets further bear the risk of spurious correlations. The more variables we can 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, e.g., linking linguistic features to economic behavior (e.g., Chen 2013; Kim, Kim & Zhou 2017; Feldmann 2019). However, without proper theoretical motivation and 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 & 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', where 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 datasets 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, e.g., the effect of topography (which only came out significant in one of four models) 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., Wassmann & Dasen 1998; Palmer et al. 2017). 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 below).

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

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

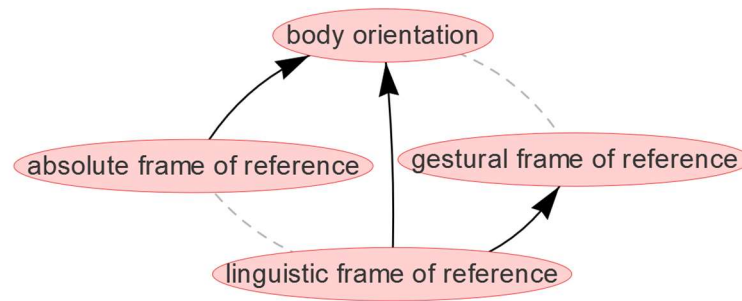


Figure 3-2: A simple causal graph based on hypothesized links between the use of an absolute Frame of Reference and body orientation during gesturing taken from Haviland (1993). The graph can be accessed in CHIELD at <https://chield.excd.org/document.html?key=haviland1993anchoring#>.

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 3-2 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 orientation³. They tested how 50 German children and 35 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 representing relationships between theorized variables and identifying links that need to be empirically tested. Additionally, once research into spatial language has been made available in CHIELD, connections between studies can be used to identify conflicts, as indicated in Figure 2.

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 & Kirby 2010).

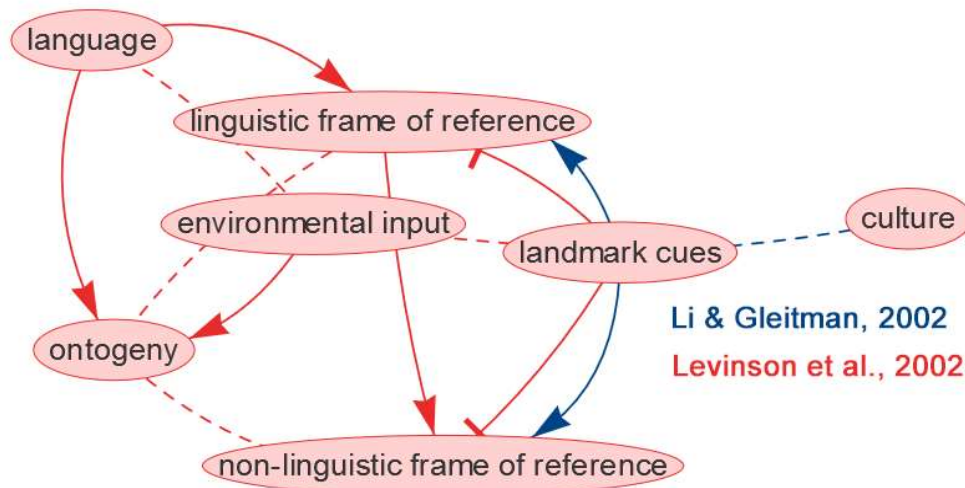


Figure 3-3: Example of a causal graph revealing a conflict in the literature (note that $X \rightarrow Y$ specifies causal links, while $X \nrightarrow Y$ denotes that X has no causal influence on Y . Colors refer to specific publications.

This causal graph (Figure 3-3) 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 & 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 Tzeltal (Polian & Bohnermeyer 2011) suggests that local salient slopes could correlate with the use of a geocentric ‘uphill/downhill’ FoR strategy, which is captured in Figure 3. 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/absence 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 & 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, Kirby & Smith 2015) or over cultural transmission (Tinits, Nölle & Hartmann 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 & Dale 2010). Another study by Nölle et al. (2020) 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 dialogue (Garrod & 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 to study spatial language on a 'large' scale: Classical tasks used to study spatial referencing and cognition cross-culturally, such as the *Man & Tree* (e.g., Levinson & Wilkins 2006), *Ball & Chair* (e.g., O'Meara & Pérez Báez 2011), or *Animals in a Row* task (e.g., Adamou & De Pascale, this issue), usually involve relations between objects in small-scale table-top 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 way-finding 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, which involves an urban lifestyle and activities such as driving, use more egocentric and less geocentric descriptions than speech-communities on the Marshall Islands (Palmer et al. 2017; Palmer et al. 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): On the one end, tasks like conversation analysis allow observing communication in its full multimodal richness, but necessarily sacrifice control over participants' behavior and communicative

conditions. On the other end, highly abstract tasks, like picture naming or reaction time experiments allow isolating 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 analogue of classical director-matcher tasks, such as the Man & Tree game, where participants explore a three-dimensional 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 (Gould et al. 2007; Bohil, Alicea & Biocca 2011) or spatial navigation research (Tarr & Warren 2002), but have only recently been introduced to the language sciences (Heyselaar, Hagoort & Segaert 2017; Peeters 2019). VR setups have great potential for studying spatial language, since they allow, e.g., controlling the environmental layout, the size or shape of objects, participants’ orientation and perspective, or modeling field-sites that are normally not easily accessible. Additionally, VR allows for manipulating otherwise realistic tasks in unprecedented ways: Nölle et al. (2020) describe OrbHunt, a collaborative VR game, where a ‘director’ has to communicate locations of ‘orbs’ to a ‘seeker’ in order to score points. This is very similar to the Maze Game (Garrod & Doherty 1994; Nölle, Fusaroli, et al. 2020) 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 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). Lastly, VR allows comparing the exact same task across different topographic environments (Figure 3-4).

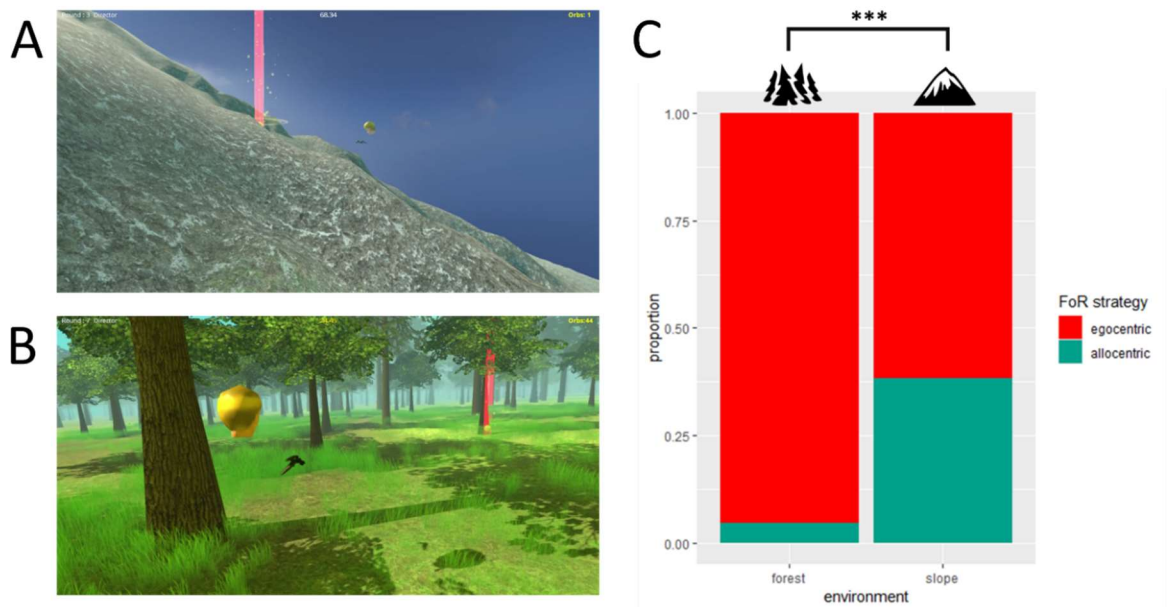


Figure 3-4: Screenshots from the OrbHunt game (Director's viewpoint) in the slope (A) or forest (B) environment. English-speaking participants were significantly less likely to purely rely on egocentric descriptions if they were playing on the mountain slope (C).

Although both environments afforded fully egocentric solutions, it was found that English-speaking participants' FoR use differed depending on whether a salient *uphill-downhill* axis was present, which afforded geocentric descriptions similar to Tseltal (Polian & Bohnemeyer 2011). This provides experimental evidence for the hypothesis that the presence of landmark cues can motivate specific linguistic FoR (compare Fig. 2 above). The next step should be to extend this approach to the emergence of artificial spatial referencing systems, to test how these are shaped by topography or cultural transmission (Kirby, Griffiths & Smith 2014).

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 large timescales in big 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 (2011; 2013) showed 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 instance, Figure 3-5 shows results of different populations evolving language in environments with different availability of geocentric or intrinsic features and its effect on the evolution of the communication systems.

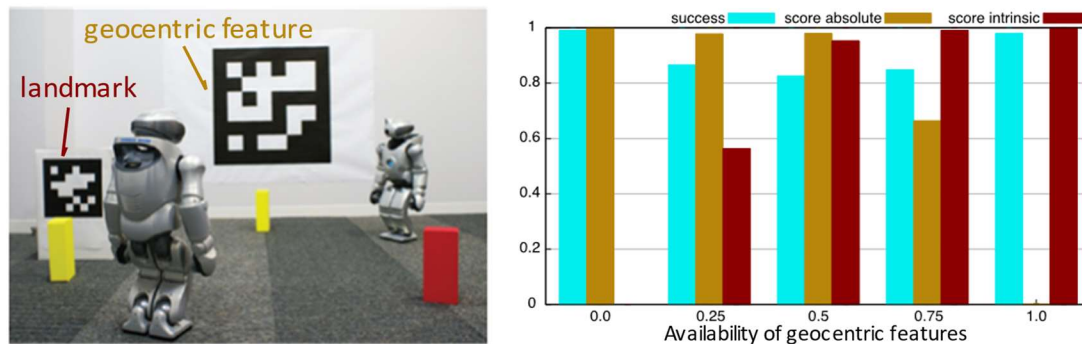


Figure 3-5: Spatial linguistic systems created by robots in interaction reflect the presence and absence of intrinsic and absolute/geocentric features in their environment. See Spranger (2016) for details.

Other agent-based models have focused on the emergence of spatial categories (Spranger 2012a), landmark-based strategies (Spranger 2012b) and spatial grammar (Spranger & 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).

4 Conclusion

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, likewise, variation in spatial referencing presents a complex 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.

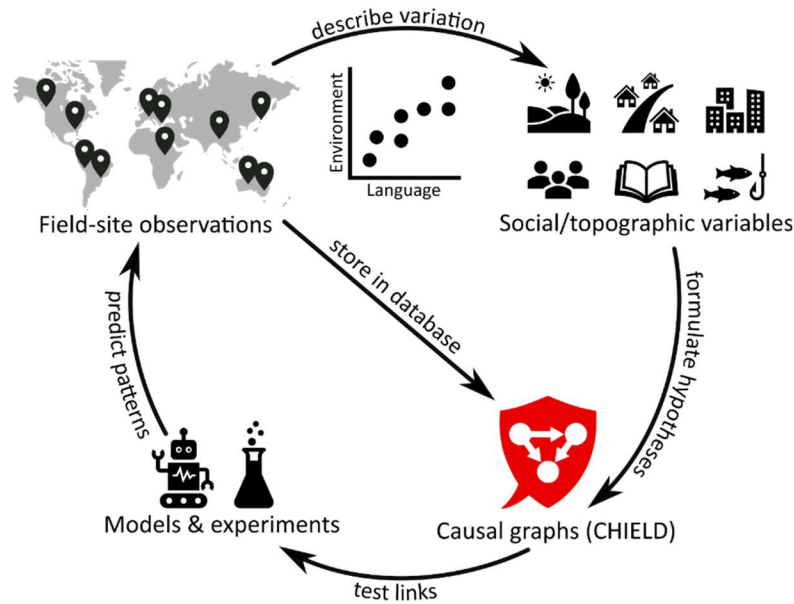


Figure 3-6: Schematic cycle using the maximum robustness approach to investigate the relationship between spatial language and socio-topographic variables.

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 (Figure 3-6).

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 easier 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 modeling the potential stages in the cultural evolution of spatial language and shed light on how socio-topographic factors shape the variation in spatial referencing systems among the world’s languages.

Acknowledgements

This paper benefitted 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, 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.

3.5 Summary of Paper 3

Paper 3 provides a framework for studying the complex interactions between spatial referencing systems and the environment from an evolutionary perspective. We suggested that combining fieldwork with experiments and computational models is necessary in order to test causal mechanisms and draw conclusions that go beyond simple correlations between linguistic and extralinguistic variables. The paper also presented the results of a virtual reality experiment that tested how native English speakers would adapt their language in a referential spatial communication game that took place in different large-scale environments. This experiment formed the first part of a series of three experiments using VR paradigms to isolate topography as a possibly variable that could affect the use of FoR strategies when describing the location of objects in large-scale space. These experiments are discussed in detail in the following chapters of this thesis.

4 Investigating the adaptivity of spatial language in virtual reality

4.1 Introduction

In this chapter, I present a series of virtual reality (VR) experiments that test the environmental adaptivity of English directly in the lab. I will first briefly review previous work that has investigated the role of the environment in grounding spatial referring expressions in controlled experiments. One limitation of research in this area has been that it is hard to create realistic settings for studying spatial referencing, which is normally embedded in complex navigation activities, in the laboratory. I will argue that immersive VR allows us, for the first time, to study spatial language in a setting that is both controlled and ecologically valid, enabling us to move from the small tabletop scale to the actual large scale, and study spatial language in vast environments that resemble the real world. I will present two experiments from a series of VR games, designed to study the impact of environment on spatial language. Experiment 1 had native English speakers solve a spatial communication game in two different environments. I found that dyads in an environment that resembled a mountainside were less likely to simply use egocentric left-right descriptions than participants playing in a forest-like environment. Instead, they would frequently make use of geocentric descriptions motivated by the uphill-downhill axis, similar to what has been described for speakers of Tseltal (Polian & Bohnermeyer 2011). Experiment 2 tested to what extent participants would switch or retain their strategy when they switched between environments. I found that in their native language that supports both FoR there is little cost to switching strategies and they would readily do so, thus providing no support for the hypothesis that the egocentric strategy is more flexible.

4.1.1 A large-scale virtual reality approach

As discussed in [Paper 3](#), the traditional method for studying how topographic environments could affect FoR usage are comparisons between different field-sites and speech-communities, where participants with varying backgrounds have to solve the same standardized tasks (e.g., Bohnermeyer et al. 2015; Palmer et al. 2017). Since the relationships between spatial language and different sociotopographic variables (such as literacy, L2, subsistence style, topography etc) are quite complex, we suggested isolating variables in controlled experiments that have been used to study the emergence and evolution of linguistic structure and conventions, such as [Paper 2](#) or the Maze Game experiment, which I also briefly discussed in the previous chapter (Nölle, Fusaroli, et al. 2020). In the Maze Game, pairs of participants had to

communicate about the location of switches (grey boxes) to help each other reach a goal position in the maze. A single environmental manipulation, namely varying the spatial configuration of the mazes, was enough to evoke different conceptualisations, which became expressed in different description schemes that participants used to communicate about locations in the mazes.

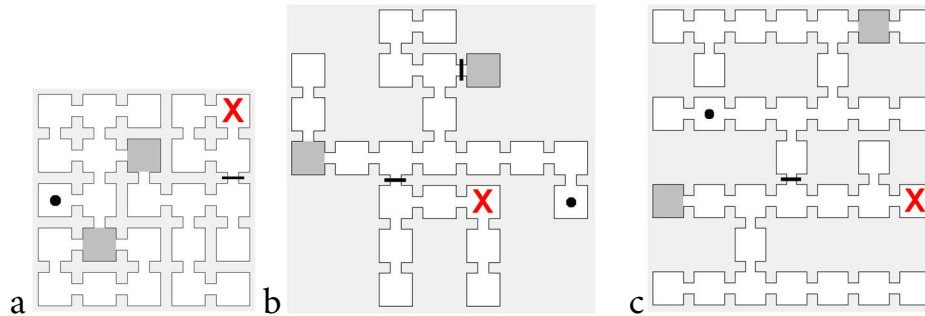


Figure 4-1: Example mazes for the regular (a), irregular (b) and stratified (c) conditions from Nölle et al.'s (2020) Maze Game experiment.

Dyads communicating about boxy mazes in the regular condition (Figure 4-1a) were more likely to routinize a description scheme relying on path descriptions (“from the top left, go one down and two to the right”) or abstract matrix-like coordinate systems (e.g., “go to b2”, based on a chess-like grid). By contrast, dyads solving mazes in the irregular condition (Figure 4-1b) were more likely to retain very concrete and figural language, using descriptions such as “it’s in the left arm, in the elbow” or “the right box in the head”. Lastly, the stratified condition (Figure 4-1c) featured salient horizontal configurations, which led many dyads to grounding a ‘line’ strategy where they would anchor location descriptions to a specific row of the maze (e.g., “first box on the second line from below” or “second box from the right on the first line from above”). In other words, dyads interacting in different environment focused on different environmental cues to construct spatial conceptualisations of the mazes, which motivated the emergence of different linguistic conventions in different environments (see Figure 4-2). A similar Maze Game study by Castillo et al. (2019) showed a similar result. In addition, they contrasted dyads with individual participants, showing that interaction is a major driver for adapting these referential conventions.

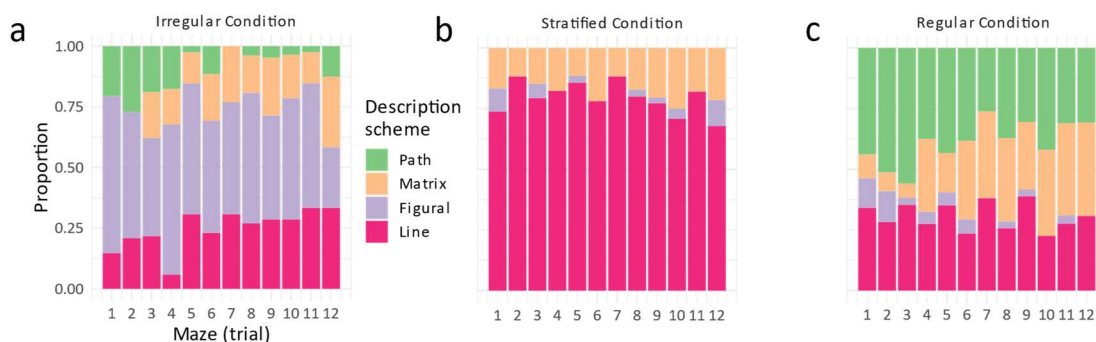


Figure 4-2: Results from Nölle et al. (2020) showing, for each environmental condition, percentages of the different description schemes that emerges over 12 subsequent trials.

In many ways, the location descriptions from the Maze Game are similar to FoR strategies that also involve locating a FIGURE object with respect to a GROUND object and involve a particular perspective or coordinate system. Participants picked up salient affordances that were in an abstract way similar to prominent topographic features and landmarks such as coastlines, slopes or rivers that geocentric FoR systems rely on in natural languages (e.g., Wassmann & Dasen 1998; Levinson 2003; Fedden & Boroditsky 2012).

However, the two-dimensional maze game seen from a birds-eye view is not suited for eliciting spatial language that would be used to describe relations in three-dimensional space as perceived from an egocentric viewpoint. For this, field-linguists would usually rely on elicitation tasks that have people describe arrays of small objects in tabletop space, such as in the Man and Tree Space Game (e.g., Levinson & Wilkins 2006: 11), or photographs of objects, such as the Ball and Chair task (O'Meara & Pérez Báez 2011). These tasks are necessarily very artificial, since they are meant to generate a standardised set of responses that can be compared across field-sites and languages. However, in the real-world spatial language is embedded in a wide variety of complex activities that involve some kind of spatial orientation. Palmer and colleagues (2017), for example, observe that the Indo-Aryan language Dhivehi spoken on non-fishing islands involves different FoR from Dhivehi spoken on islands where fishing is the predominant mode of subsistence. The use of the geocentric FoR that relies on topographic features of the atoll environment ('lagoonward' and 'oceanward') is directly embedded in cultural activities that engage with the sea. Generally, a central aspect of both human and animal spatial cognition is navigating large three-dimensional environments and creating mental maps to find directions (Hutchins 1995; Allen 1999; Thiering 2014). Levinson (2003: 216) argues that spatial language is deeply interlinked with spatial navigation and wayfinding, and cultural practices related to these activities and their variation are "deeply linked to specializations in frames of reference in language". It is therefore surprising that very little navigation activity has actually been studied by anthropologists, or as Levinson (2003: 217) puts it:

“Still, I think it must be conceded that in many ways we know much less about navigation in our own species than amongst birds, bees and ants. Apart from the efforts of the geographers, there are simply relatively few studies of how humans actually find their way around real novel environments, or calculate angle and distance and current location in moving around on a scale larger than the psychological laboratory. One might have expected anthropologists to have had a keen interest in wayfinding amongst, especially, hunter-gatherer groups. But, on the whole, the information available is extremely disappointing.”

One problem with complex cultural practices is that they involve many different variables. As such, it is hard to study them in a controlled setting. Traditionally, if ecological validity is desired, more qualitative ways for analysing data are used to describe the finer details of the phenomenon that is studied. In anthropology, this could be ethnographic fieldnotes that serve as collections of descriptive and reflective information through which a qualitative picture of the object of study emerges. For the study of language, an example would be conversational analysis, the study of open-ended natural dialogue, where even the finer details of prosody, gestures, facial expressions can be annotated to study language qualitatively in its full multimodal richness (Holler & Levinson 2019; Hutchby & Wooffitt 2008).

This is in sharp contrast to experiments that are used to study these components of communication in isolation in order to allow researchers to draw inferences about the underlying processes. To ensure experimental control, tasks usually involve unimodal stimuli that are presented out of context and participants have to engage in repeated activities that allow measuring a clear response variable such as reaction time in a picture-naming or sentence-processing task. Peeters (2019) has pointed out that ecological validity and experimental control are often seen as the ends of a continuum, where one has to be sacrificed in order to increase the other (Figure 4-3a). However, Peeters argues that this continuum is misleading, since it is theoretically possible to design tasks that combine a high ecological validity with tight experimental control. This is possible with new technology that allows immersive VR tasks that simulate communication in its natural mode – multimodal face-to-face interaction – while allowing to control variables of interest in the same way as traditional artificial tasks used in psycholinguistics. Immersing participants within a virtual world which the experimenter can control meticulously allows simulating day-to-day language use in its full richness, while making sure that the variables of interest can be measured under controlled conditions (Figure 4-3b).

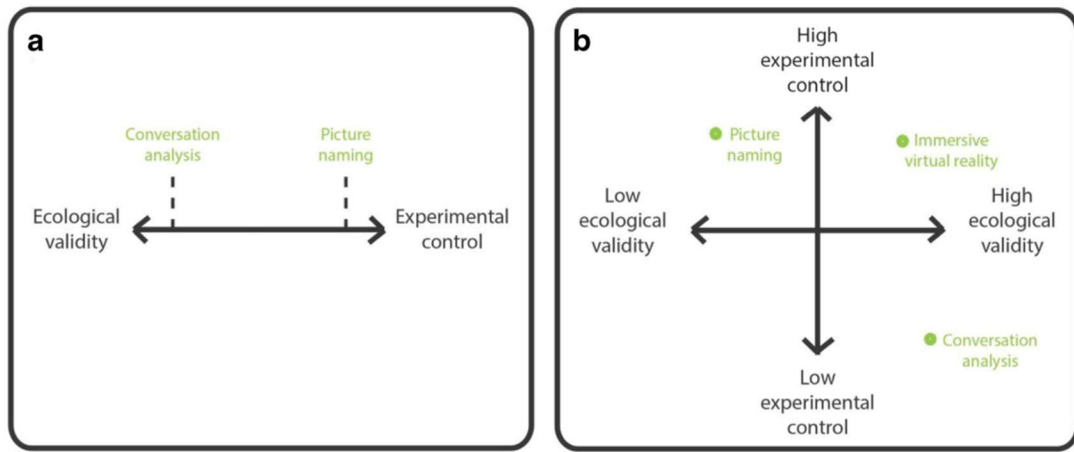


Figure 4-3: Ecological validity and experimental control in the language sciences perceived as two extremes on a continuum (a) or as two orthogonal factors (b). Figure taken from Peeters (2019).

While other research fields, e.g., in rehabilitation are already using VR paradigms extensively (e.g., Howard 2017), this technology is only now being adopted by the language sciences (see Peeters 2019 for a review of pioneering research).

For spatial language, VR bears a huge potential, since it allows us to study the use of spatial language for the first time in controlled large-scale settings without the need to rely on photographs or objects in tabletop space that do not necessarily reflect how spatial language would be used on a larger scale. Instead, we can look at the impact of environments that appear just like those that have been hypothesized to affect languages in the real world, such as the presence of salient topographic landmarks.

A first study that went into this direction was the Virtual Atoll Task (VAT) by Lum and Schlossberg (2014) that was used to elicit spatial descriptions in Dhivehi and Marshallese speakers. The task involves a director and a matcher playing a collaborative game in front of a laptop computer. The VAT is designed in the style of a third-person perspective videogame, where the matcher controls a boat and has to navigate a virtual atoll environment in order to find treasure chests (Figure 4-4). The director knows the location of the treasure chests and can guide the matcher using language.



Figure 4-4: Screenshot of the VAT from Lum and Schlossberg (2014: 84).

While this task was successful in eliciting spatial language related to a realistic navigation setting that mimic ocean-voyaging, it did not yet make use of immersive VR in the sense of actually placing participants *in* the scene. The authors acknowledge themselves that the use of computers could be associated with left-right systems and thus bias speakers. There were also some issues with inexperienced computer-users struggling with the controls.

In the following, I will present a novel paradigm that takes this approach to the next level by fully immersing participants inside realistic virtual 3D environments, where they have to solve collaborative spatial communication games. The first experiment that is described below, tested to what extent the spatial language participants would use is sensitive to the local topography of these virtual environments, i.e. whether large-scale topographic features affect the FoR participants use to express spatial relations in the game.

4.2 Experiment 1: FoR adaptivity in English

4.2.1 Outline and prediction

I built a VR paradigm that allowed me to immerse participants in realistic large-scale environments that systematically varied in terms of their topographic affordances. The idea was that linguistic routines that are established by a dyad for talking about spatial locations in these environments would reflect salient affordances present in these environments. ‘Affordances’ here means salient invariant physical properties that relate to possible actions in the environment (Gibson 1979). In this way, the presence or absence of

specific topographic features like a slope or absolute landmarks in the environment was hypothesized to affect how participants reason about the task in a similar way to the different maze figurations discussed above. In this way, the experiment can serve as a simulation of how speakers of the same language solve spatial referential communication problems at different field-sites with different topographic features. As discussed in [Paper 3](#), for real languages, this interaction of spatial language with the environment is further complicated by cultural and social factors (e.g., Bohnemeyer et al. 2015; Palmer et al. 2017). The advantage of an experimental approach lies therefore in isolating one specific variable to test its impact on the linguistic system.

The task was a simple coordination game in which players had to communicate about the location of objects, thus eliciting spatial location descriptions similar to previous studies utilising collaborative games (Garrod & Anderson 1987; Lum & Schlossberg 2014; Castillo, Smith & Branigan 2019; Nölle, Fusaroli, et al. 2020) and the classical table-top director-matcher tasks used to study spatial typology (see, e.g., Levinson & Wilkins 2006). These descriptions could then be studied with regard to their underlying FoR strategies. I predicted that participant pairs communicating in English would predominantly rely on an egocentric FoR strategy if the task would take place in a forest-like environment that did not feature salient landmarks. By contrast, I expected pairs playing the same game in an environment resembling a mountain slope to use proportionally less egocentric language and incorporate more descriptions that rely on an allocentric/geocentric FoR afforded by the environment. Such a difference in FoR strategies between environments in an otherwise identical task would provide experimental evidence that the use of FoR is to some extent sensitive to topographic features of the environment at hand.

4.2.2 Methods

Participants

42 native English speakers ($N = 21$ dyads) were recruited via Edinburgh University's MyCareerHub service and received £5 in compensation for taking part in the study. Participants were screened for normal or corrected-to-normal vision and advised to not take part in the experiment if they had a tendency to get motion sick or suffered from vertigo, to make sure that they would not be negatively affected by the VR. Two additional dyads were excluded due to technical issues that occurred during data collection.

Materials

The experiment took place in the Centre for Language Evolution's open plan lab at the University of Edinburgh. The room was cleared, to provide approximately a 2×2 m play area, where two participants at a time could be placed in diagonally opposing corners, so that they would be surrounded by enough empty space to be able to fully extend their arms without colliding with each other or the interior. The lab was equipped with two HTC Vive headsets including two sets of controllers per participant, and two VR-ready HP computers running connected to the network and running the setup for each participant (see [Appendix B](#) for full technical specifications). The Lighthouse base stations were mounted on poles in diagonally opposite corners of the room, so that each player's movement could be tracked by at least one base station at any given time.

OrbHunt: design and objective

To simulate the adaptation of spatial language to a particular environment, a two-player collaborative VR game called 'OrbHunt' was designed to evoke spatial linguistic descriptions in one of two environments. The objective of the game was to score points by collecting 'orbs', yellow spheres placed in the environment. At the beginning of each experimental session, both members of a pair were assigned two complementary roles, the 'Seeker' and the 'Director'. Only the Seeker could collect orbs, and thus score points. This was made more difficult by a feature programmed into the orbs: An orb would only be visible to the Seeker, if they came sufficiently close to it, that is if the Seeker entered a spherical area with a radius of 5m surrounding the orb. Thus, randomly stumbling upon an orb was unlikely, particularly because players had a maximum of two minutes per round to find as many orbs as possible.

By contrast, the second player, the Director, could not collect orbs themselves but could see them from any distance. To the Director, orbs appeared highlighted with glowing light effects and a transparent red light pillar indicating their location even from far distances (Figure 4-5). Similar to the Maze Game (Garrod & Anderson 1987; Nölle, Fusaroli, et al. 2020), the Director could thus help the Seeker reach and collect the orb by communicating its location using spatial descriptions.



Figure 4-5: Screenshot from the OrbHunt game, showing the Seeker's avatar (left) and an orb from the Director's point of view (right) that would appear invisible to the Seeker.

Note that in the real-world, such a task could easily be solved with deictic methods. The Director could simply point into the direction of the orb or move to its location. Since the experiment aimed at eliciting spatial location descriptions, an additional manipulation was added: The Director could see the Seeker, but not vice versa. Similar to the orb's area of visibility, such manipulations are a significant advantage of VR paradigms. In the real world it would be extremely difficult to introduce such perceptual asymmetries, but in VR, the Director's avatar can simply be rendered invisible to the Seeker's in-game camera. Via this manipulation, deictic strategies were effectively disabled, and spatial linguistic descriptions constituted the main resource for helping the Seeker collect orbs and score points.

At the beginning of each round both players and a first orb would spawn in random locations near each other. The Seeker could then try finding the orb by coordinating with the Director. Every time an orb was collected, another orb would spawn in a random location until the time was up. After each round both players would switch roles, the previous Seeker becoming the Director and vice versa, for a total of 10 rounds. This combination of spontaneous dialogue and the participation of both players in the game allowed simulating a routinisation process, where players could over time develop a spatial referencing strategy through precedence, feedback/clarification and alignment, similar to related joint problem-solving tasks (Healey & Mills 2006; Healey 2008; Garrod & Pickering 2009; Fusaroli & Tylén 2016) and in relation to the environment at hand (Nölle, Fusaroli, et al. 2020).

In-game avatars

To provide both players with virtual bodies inside the VR, I created unisex avatar heads in Blender (Community 2018), which were then imported into the Unity 3D game engine. The heads track the VR headset position, while two realistic 3D models of the Vive controllers track the movements of their real-world counterparts. The avatars were thus not full body representations, but more akin to floating heads and hands that corresponded to their orientations in the real world (Figure 4-6). Note that the Vive headset requires users to look directly ahead into a relatively narrow visual field to receive a sharp image. Players therefore must turn their head to look around. Due to this, the forward vectors of the avatar heads reliably mirror players' looking directions. This enables the Director to take the Seeker's egocentric perspective for location descriptions, and it also allowed gathering data about players' looking directions (which is particularly relevant to experiment 3 discussed in [Chapter 5](#)). Avatars only differed in hair colour and different attributes attached to their game objects in Unity, which made it possible to control what each player could see.



Figure 4-6: Screenshot showing avatars for the Seeker and Director rendered in the Unity 3D game engine.

Game controls

Since space in the lab was limited and could not simulate unrestricted movement in vast open or topographically complex environments featuring elevation, a point-and-teleport system was implemented. All interactive elements of the game relied on the controllers of the HTC Vive and were built using the open-source Steam VR plugin for Unity (see Murray 2017 for an introduction).

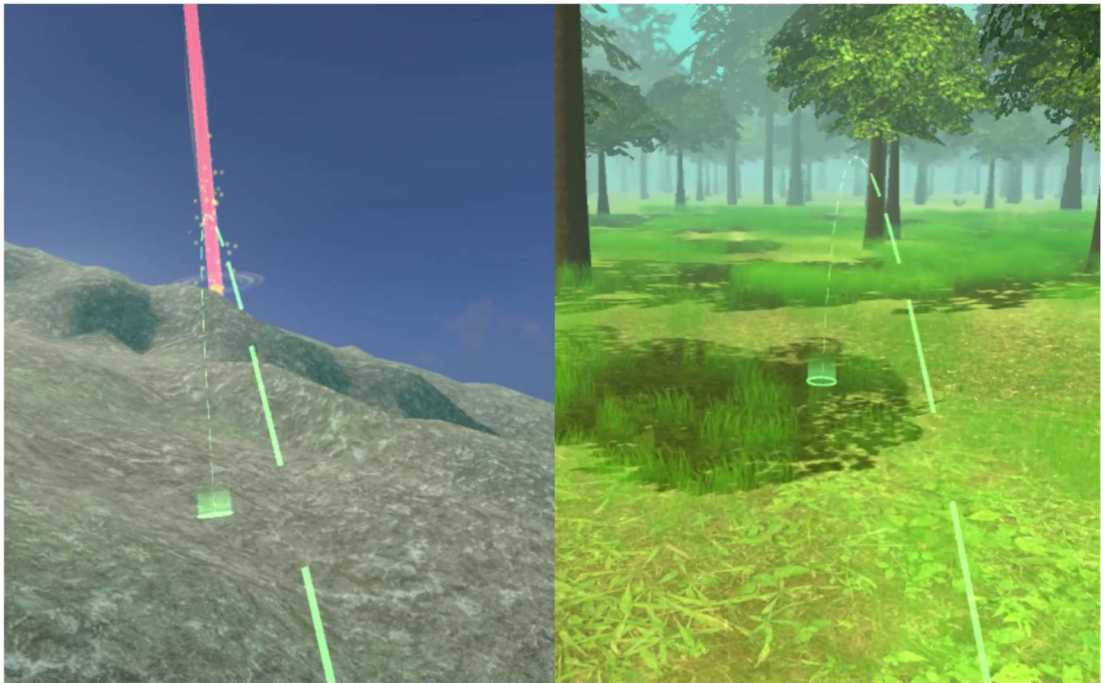


Figure 4-7: Screenshots showing the pointer arc that allowed players to teleport their avatar to a desired location in the game. The pointer showed green for all locations the players could move to and turned red if players pointed at invalid locations, e.g., invisible walls fencing off the outer areas of the map.

Players could press the trackpad on their controllers (Figure 4-7), after which a pointer would appear that allowed them to almost instantaneously move to a desired location within a maximum distance of approximately 10m (Figure 4-7). While this way of locomotion is unnatural, it is quite easily learnt and has several clear advantages over other forms of movement. Firstly, it allowed implementing a stationary setup, where each player only required a 1×1 m square play area, in which they could freely rotate and move. Secondly, not having to walk made the game much less tiring. Finally, the lack of translational movement is extremely effective in preventing participants from experiencing motion sickness within VR worlds (Bozgeyikli et al. 2016; Rahimi, Banigan & Ragan 2020). The only known caveat of teleportation systems is that they can lead to a slight overestimate of traversed distances (Li et al. 2018), which I did not expect to interfere with the task.

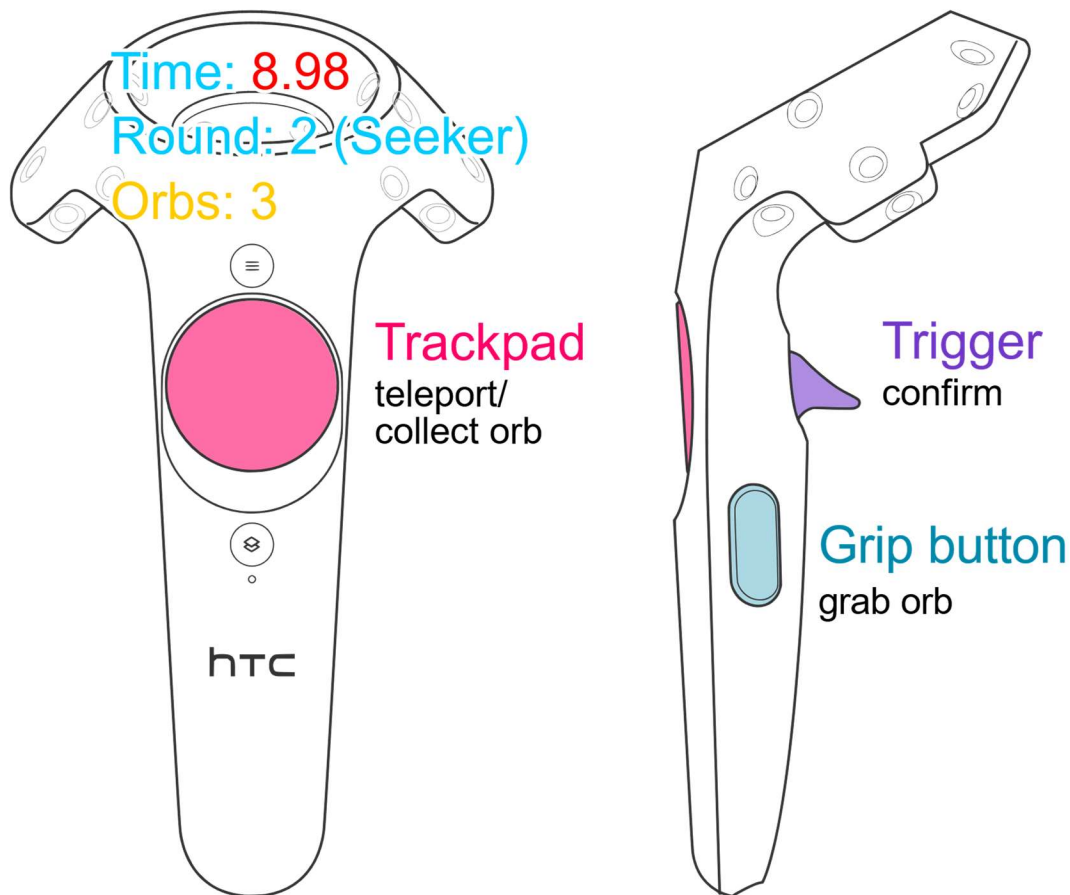


Figure 4-8: Schematic diagram of a game controller showing the button layout. Within the VR game one controller featured a two-dimensional canvas constantly showing the current round time, round number, player role and score. Note that the button assignment changed for experiment 3 (see [chapter 5](#)).

Once an orb was found, the Seeker could collect it by grabbing it using the grip button, which would attach the orb to the controller, and then proceed to collect it using the trackpad (Figure 4-8). The orb would then disappear, increasing the score by one and spawning a new orb in a random location within a circular area ($r = 35$ m) around the position where the last orb was collected. This was done to make the orb locations as random as possible within the environment. If time ran out at the end of a round, the orb would disappear and could not be collected any longer. For each player, one controller also had a virtual canvas attached to it that informed them of the current round number, score, current role and the remaining time. The timer would turn from white to yellow when 30 seconds were left and to red when 10 seconds were left (Figure 4-8).

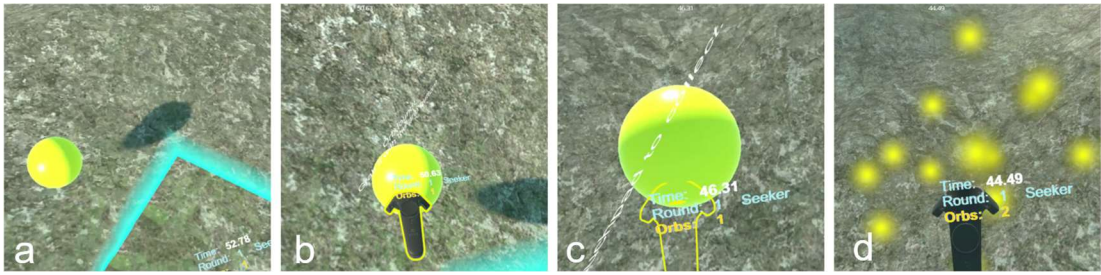


Figure 4-9: Screenshots showing the Seeker collect an orb. a: The Seeker spots an orb on the ground near their play area (blue square that assured that they would not collide with objects or walls in the real world). b: Upon ‘touching’ the orb, the controller lights up and a message appears on the orb stating: “Orb found! Grab it with the grip button!”. c: The orb is attached to the controller and can be collected using the trackpad. The message changes to “Press touchpad to collect”. d: The orb disappears with a particle animation upon collection and the score display is incremented by one.

If the end of a round was reached, a message would appear in each of the players’ visual field, stating the role they would assume on the next round, e.g., “Next round you will be the director. Pull the trigger to continue”. Players could then use the trigger button on any controller to confirm and proceed to the next round.

Virtual environments

The game took place in one of two environmental conditions designed to afford specific spatial referencing strategies. The **forest** consisted of a flat terrain with a square area of 500×500 m populated with hundreds of trees. There were four spawn points distributed around the centre of the map where a round could start, and the traversable area was constrained by invisible walls, so that players could not accidentally reach the edges of the map (Figure 4-10).

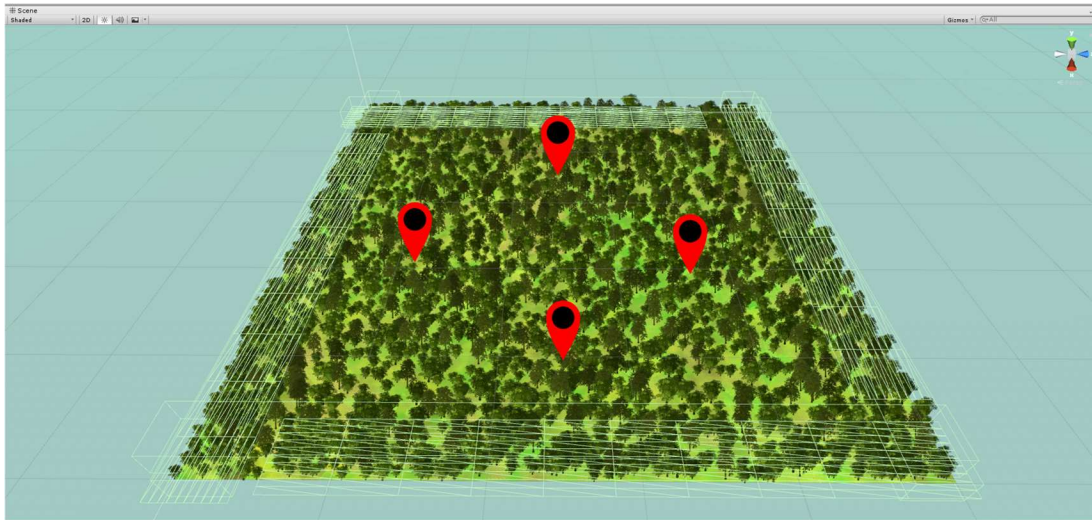


Figure 4-10: Aerial view of the forest environment showing four spawn locations as well as the invisible walls (green wireframes) surrounding the traversable area.

Trees varied in size, shape and rotation, and the environment featured different grass and ground textures, a wind effect moving grass blades and tree branches, and directional light casting real-time shadows. A fog effect was used to constrain the visual range and make the forest appear infinite in every direction (Figure 4-11).



Figure 4-11: In-game view of the forest showing textured trees, grass and the fog effect.

Importantly, the skybox featured only a cloudy texture and did not contain a distinct visible sun that could be used as an absolute anchor. This created a dense environment, where it was easy to lose a sense of absolute directions, since no particular salient landmark stood out. Hence, I hypothesized that the forest would provide a strong affordance for the egocentric FoR to talk about locations and directions.

By contrast, the **slope** environment featured a topography that was akin to a mountainside. The terrain was designed in Unity on a 1000 × 1000 m plane, which was conically raised towards the centre to create the impression of a large mountain. I used several rock textures to simulate an increase in ruggedness and decrease of vegetation as a function of height. Algorithms from the Erosion Brush plugin (Pahunov 2015) were used to create a surface topography that appeared to have been shaped by realistic hydraulic action, weathering and sediment. The playable area was fenced off by three intersecting invisible walls, which contained four spawn points where a round could start (Figure 4-12). This prevented players from going ‘around’ the mountain, which caused the uphill-downhill axis to appear constant and absolute during the game. The skybox featured a much clearer sky texture and included a sun and moon that appeared opposite of each other and perpendicular to the uphill-downhill axis. This was done to additionally enable the grounding of cardinal-like systems with four orthogonal directions (uphill, downhill, sunward, moonward), since the ‘across’ axis would otherwise remain ambiguous.



Figure 4-12: Aerial view of the slope environment showing four spawn locations as well as the invisible walls (green wireframes) surrounding the traversable area.

Similar to the forest, a fog effect was used to make the lower area fade out towards the horizon and conceal that the terrain was finite. While it is hard to convey a sense of space in two-dimensional screenshots of the game, participants reported that they felt very present in the game world and it did feel like they were looking up or down a realistic mountain.

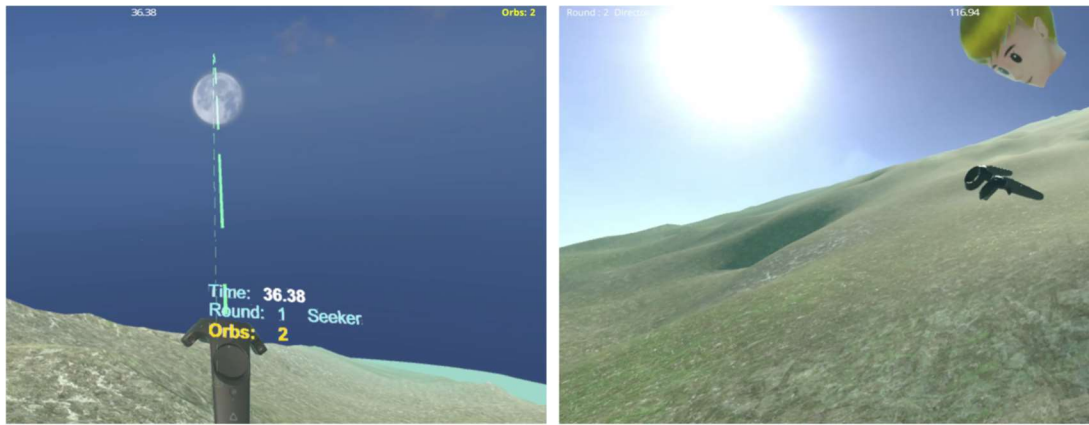


Figure 4-13: In-game view of the slope environment. The moon (left screenshot) and sun (right screenshot) were aligned with the ‘across’ axis perpendicular to the traversable slope.

The slope was designed to evoke a strong sense of allocentric directions based on the slope and celestial bodies. It was thus hypothesized that the slope would strongly afford descriptions relying on the geocentric FoR, similar to what has been documented for Tseltal speakers at different field-sites that feature topographically salient slopes (Brown & Levinson 1993; Polian & Bohnemeyer 2011).

In sum, the two environments were designed to provide strong affordances for the egocentric FoR (forest) and geocentric FoR (slope), respectively. Furthermore, both environments did not include orientable objects with facets that could be referred to as *fronts*, *backs* etc. This was done to be able to focus on the contrast between egocentric and geocentric without providing an additional affordance for intrinsic solutions. In this way, the setting in which the task takes place, and the referents present in its environment, can already constrain the space of possible spatial referencing strategies.

Procedure

Participants entered the lab, and after providing written consent, were assigned to their play area by the lead experimenter, who would explain the controls of the game and show the buttons on the physical controllers. Before the start of the first round, participants were placed in a simple training environment. This environment was a featureless flat plane with a simple grass texture and empty blue sky, to not prime them in any way with landmarks or directions. Participants could then try out the point-and-teleport mechanic and familiarise themselves with how to move around while remaining inside the safe play area indicated by a blue square on the ground. They were instructed to always remain in the blue square to prevent colliding with anything in the room, whereas within the square they could turn around freely.

During the experiment, two experimenters assured that participants could not entangle themselves in the headset wires when they rotated their bodies and that they did not step out of their designated play area⁴.

After participants confirmed that they could see the in-game models of their controllers and were able to teleport, the lead experimenter would explain the rules of the OrbHunt game (see instructions in [Appendix C](#)). Each dyad was assigned to either the Slope or the Forest environment, where they would play 10 rounds of the OrbHunt game. Depending on which condition they had been assigned to, each dyad was then transported to either the slope or forest environment. Before starting the game, the experimenter explained that the players could now check on their controllers to see the time, score, and role displays, and that once the first round started, the controller would indicate which role they would start out with. The experimenter would also repeat that the Director should then be able to see the Seeker and an orb somewhere in the environment, so that they could begin working together to collect it.

The experimenter then started the game manually by pressing a key. At the beginning of every round, the Seeker would spawn randomly in one of the four spawn positions in the environment. The Director would spawn in a random position within 5 m of the Seeker and the first orb would spawn within 35 m of the Seeker. Each dyad played 10 rounds of 120 seconds. After each round the roles switched (the Seeker from round 1 would become the Director in round 2 etc). At the end of each round, when the timer reached 0:00, any orb in the environment would disappear and both players had to confirm by clicking with the controller to start the next round. Once both players confirmed, the new round would automatically start at a random spawn point. At the end of the final round, a message was displayed that informed participants that the experiment was finished and that they could remove their headset.

For all games, both video data (from the perspective of each player's in-game cameras) and audio data (the spoken dialogue recorded via microphones attached to the headsets) were recorded. The game was programmed to automatically record the positions of both players in 3D space, their looking direction, and the position of the current orb whenever important events took place (teleporting, collecting an orb, confirming to start the next round etc), and wrote them into a CSV file. Additionally, this data was recorded at 10 frames per second in a separate CSV file allowing to reconstruct where players had been looking or moving at any given moment during the game.

⁴ I would like to thank Paola Peña, Elliot Huggett, Jasmine Salmon and Lucie Wolters for assisting me as experimenters in the lab.

4.2.3 Results

Performance

Overall, all pairs were able to solve the OrbHunt game and collected between 21 and 58 orbs. The median score was 39 orbs in both environmental conditions (10 dyads playing in the forest, and 11 dyads playing on the slope). This suggests that the task difficulty was very similar, and the environment where it took place did not affect performance (Figure 4-14).

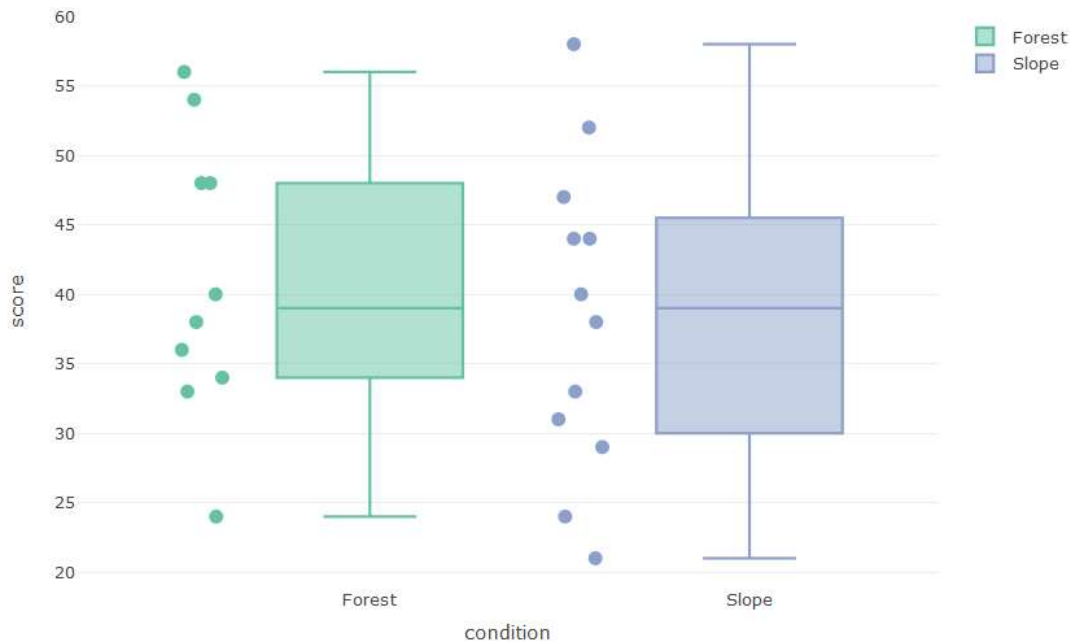


Figure 4-14: Number of collected orbs for both environmental conditions. Dots represent scores for individual dyads. Lines of the boxplots indicate median scores per condition. The box represents the interquartile range, and whiskers represent the range of scores.

Coding of the video material

Each experimental session resulted in two video files – one for each player – recording everything that happened during a session from the players' perspectives as well as their spoken dialogue. All videos were then transcribed using verbatim transcription. Table 4-1 shows an excerpt of the resulting transcript for pair 01 who were playing the slope condition:

Table 4-1: Example of a dialogue from the start of a game until the Seeker successfully found the first orb (pair 01 in the slope environment).

p01 Who's what?
p02 Yeah I'm the director. So where is she, I can't see her
p01 I'm on the top of a mountain

p02 ok, I can see the golden orb
p01 ok, you need to come down.
p02 Got it coming
p01 Good
p02 Is that ok?
p01 Nono you need to move one
p01 ..down
p01 come a little up.
p02 A little up
p01 Ok uhm. Just move to your right, good. Good move to your right
p02 Is that ok?
p01 Move a little more.
p02 Oh found it.
p01 Yeah
p02 I need to go through the-
p01 Yeah go a little ahead. Erm I have a doubt. I'm just supposed to stand right? I don't need to go collect it?
p02 How do I collect it. I keep. I'm very close actually. Oh got it.
p01 I think she got it.
p02 I got it now what do I do? Got it. Next?

Each dialogue segment containing a spatial referential description was then treated as an individual utterance (i.e. row in the dataset) and coded for its underlying spatial referencing strategy. The full dataset containing 10578 utterances across 21 dyads is available at https://osf.io/4azur/?view_only=3d977abf51964087a1aa61c2f28dd233. In total there were 16 categories that an utterance could be classified as belonging to. These categories are presented in more detail in the following section.

Spectrum of referential strategies

Frames of Reference

Firstly, in order to study which Frames of Reference were employed by participants playing the OrbHunt game, utterances were coded with respect to the detailed FoR classification proposed by Bohnemeyer et al. (2015) discussed in chapter 3 (Table 3-1). This resulted in six FoR categories: egocentric extrinsic, egocentric intrinsic, allocentric intrinsic, landmark-based, geomorphic, and absolute.

The egocentric extrinsic or 'relative' FoR could for instance be used in the forest environment to locate an orb (the figure) next to a tree (the ground) relative to the Seeker's viewpoint (the anchor), such as in (1).

- (1) p41: the orb is right of the tree

In English, the intrinsic axes of the observer are translated to the ground object (the tree) to project their body's right onto the tree. By contrast the egocentric intrinsic FoR conflates anchor and ground by simply extending the intrinsic axes of the observer and setting the figure directly in relation to that:

(2) p04: It's, it's just on your right

This 'direct' frame (Danziger 2010) was very common in the game, since the OrbHunt task focuses on the Seeker locating the position of the orb. Therefore, the Director would often take the Seeker's perspective as an anchor and set the orb directly in relation to that. Since the task was about navigating an environment it was also common to combine such directions with a 'turn' command, like in (3):

(3) p06: and now turn to your right

The allocentric intrinsic or object-centred FoR is usually reported to be used frequently by speakers of English and related languages like Dutch or German (Carroll 1993; Danziger & Pederson 1998; Levinson 2003; Van Staden, Bowerman & Verhelst 2006). However, experiment 1 was designed to investigate whether the topographic environment could have an impact on the preference for the egocentric or geocentric FoR. Therefore, environments were designed to not contain objects with clear functional facets that could be used to anchor intrinsic relations (such as houses, cars, chairs that possess *fronts* and *backs* etc). As a result, the corpus of dialogues did thus not contain a single utterance expressing an object-centred spatial relation.

As far as the geocentric FoR is concerned, landmark-based descriptions were common in both environmental conditions: On the slope, participants would often refer to salient topographic features, such as jags, dips, or the top of the hill that were turned into ad hoc landmarks (4-8)⁵:

(4) p04: go to the top

(5) p06: get to the like where the dip goes ..

(6) p39: just follow the jag ..

(7) p39: go up that tooth

(8) p40: ok, and now, go, so now go on the high part of it

⁵ It should be noted that in contrast to conventionalised landmark-based FoR systems, these descriptions remained *ad hoc* and only use landmarks as 'topological' cues (see also Piaget & Inhelder 1956) or what Mishra et al. (2003: 369) call a 'situationally based local landmark' in contrast to fully conventionalised landmark-based coordinate system such as in *the ball is mountainward of the chair* as, e.g., in Balinese (see Mishra, Dasen & Niraula 2003).

In these cases, the directions correspond to vectors towards an environmental entity that serves as an anchor to locate the figure (the orb) with respect to the ground (the Seeker). This was also possible in the forest environment, where features such as open areas, flowers or specific trees could be used as an anchoring landmark:

- (9) p26: yeah, go straight through the flowers
- (10) p20: so go towards that open space
- (11) p19: so go back to that tree

However, the geomorphic variant of the geocentric FoR was only afforded on the slope, where the *uphill*, *downhill* and *across* axes could be abstracted and used readily at any given position in the game (12-15):

- (12) p09: ok, it's reappeared, slightly uphill above you
p09: if you go straight up the hill
p09: ok uhm, now it's back downhill where you started
- (13) p15: ok now, try and go across the slope
- (14) p30: ok, so you wanna go straight across
p30: nonono turn back to your, like facing across the hill
- (15) p06: and go straight along the hill
p06: don't go up

In natural languages, such geomorphic directions can be further abstracted to encode fixed absolute bearings that can be used in a similar fashion to the cardinal directions *north*, *south*, *east*, *west* in English (e.g., Levinson 1997; Levinson 2003; Wassmann & Dasen 1998; Polian & Bohmeyer 2011). Occasionally, participants attempted to directly use the cardinal directions with very limited success. For instance, on two separate occasions, participant 40 tried to use the direction *northwest* in the forest environment, but their partner was unable to understand what they meant. Since there was no cardinal reference point or absolute landmark, they were unable to ground such a cardinal system, as the Seeker could not infer which direction corresponded to *north* (16-17).

- (16) p40: go, northwest
p39: that doesn't help
p39: there's no sun in the sky
- (17) p40: now, from this position, go northwest
p39: north- doesn't-
p40: it doesn't make any sense?

p39: no
p39: just tell me straight

Another case was participant 09 combining geomorphic *uphill* with *northwest* to describe a diagonal axis across the mountain slope, but their partner did not seem to understand (18):

(18) p09: if going uphill...
p09: ... was north, go northwest
p10: uh?
p09: uhm, no, so 90 degrees

The idea was to equate *uphill* with *north* in order to use cardinal terms for more precision. However, since player 10 did not understand what was meant, the pair did not revisit this strategy and simply kept combining *uphill/downhill* with a degree-based systems indicating how far the Seeker should turn away from this axis (e.g., *downhill and 45 degrees to the left*).

Lastly, one pair used the cardinal directions in an egocentric, 'field-based' sense (see also the following section), where *forward = north*, *left = west*, *right = east* and *backward = south* (19):

(19) p23: ah, and then further north
p24: so, is that forward?
p23: yeah, forward

This example therefore does not actually rely on an absolute FoR and instead simply exapts a cardinal term to express an egocentric intrinsic direction. This is very similar to other field-based systems based on a clock or angles that will be further discussed below.

Taken together, all FoR strategies except the intrinsic and absolute frames were used by pairs playing the OrbHunt game and occurred in the dialogues. In all of the examples above, it is also clear that, since the objective was for the Seeker to collect orbs, an orb served usually as the figure object, and the Seeker (or their position) as the ground object. Since this was contextually established, many descriptions could underspecify figure, ground or both. For example, *left* could mean *the orb is to your left* in the egocentric intrinsic sense. The focus on giving directions based on the Seeker's looking direction is to some extent a result of the design of the game, which I will discuss further below.

Other referential strategies

Beyond Frames of Reference, the OrbHunt game also afforded other referential strategies. A very typical strategy that was already briefly mentioned above relied on combining field-based systems with an FoR strategy. Talmy (2000) distinguishes field-based systems from Levinson's (2003: 47–50) absolute systems in that they can involve various different secondary reference objects that encompass a primary reference

object. This includes not only the earth-based field as in *The bike is east of the church*, but also other reference objects, such as the queue in *John is ahead of Mary in line* (Talmy 2000: 212). Field-based systems, in contrast to Levinson's definition of the absolute FoR, can therefore be relative as well as absolute (Talmy gives the example of an astronomer considering earth-based compass points with respect to celestial orientation). In the OrbHunt game, the most frequently used fields were a circular angle system and a clock. For example, circular degrees could be combined with *left/right* to be more precise about the direction of the orb:

(20) p09: so turn about 135 degrees to your left

In such a relative circular field, 0° corresponds to the front, i.e. the current looking direction of the Seeker, and 180° corresponds to the opposite vector, i.e. *behind* from the Seeker's perspective. The use of relative terms like *left* or *right* (from the Seeker's egocentric perspective) indicates whether the rotation from 0° is clockwise or counterclockwise. Such a strategy allows defining the direction more precisely in situations where the director does not want the Seeker to simply turn 90° to their left or right.

Interestingly, in some cases the circle could also be aligned with geocentric coordinates:

(21) p43: ok, uh, so if you use the top of the hill as like a reference

p43: and then turn 90 degrees to your right

In (21), the Director uses the geocentric *uphill* direction as a starting point for the rotation, and then specifies the direction of rotation using the egocentric description *90 degrees to your right* (i.e. clockwise).

Similarly, several pairs made use of clock positions. In this system, the reference field is a horizontal 12-hour clock face with the front at 12:00 (i.e. the Seeker's egocentric front). In such a system, *3 o'clock* means *to the right*, *6 o'clock* means *behind*, *9 o'clock* means *to the left* and so on. The main difference to the degree-based system is that the clock corresponds to a full circle, so left and right do not have to be specified to communicate the direction of rotation⁶:

(22) p19: so, it'll be the third tree, like, third tree, at 10 o'clock

(23) p26: uhm, turn to about your 5 o'clock, 4 o'clock- almost all the way around

(24) p46: turn- uh 11, 11.5 o'clock- huh

⁶ A bearings system would use degrees in this way.

In (24), the Director attempts to give even more precise directions by further segmenting the hour between 11 and 12 to indicate that the orb is located just ever so slightly to the left of the Seeker. In this way, the clock system could be used to describe a large number of different angles in a similar fashion to a degree-based system.

These FoR-based strategies were supplemented by other recurrent description types that were observed in several pairs and thus coded as separate utterance categories.

One frequently used description type was for the Director to express something along the lines of *forward/ahead/further/keep going* to communicate that the Seeker was going in the correct direction, but had to keep teleporting further in that direction to find the orb (25-27). Other common commands had the Seeker stop and wait for further directions (28-29), or simply turn around and go into the opposite direction (30-31):

(25) p01: a little more

(26) p01: come on forward

(27) p01: good, keep going

(28) p10: erm stop there, oh wait

(29) p17: er, wait, stop, from where you are

(30) p09: ok. turn the other way

(31) p13: ok so, uhm, turn around

Often players would also try to indicate the distance by using terms like *near, far, short, long* (32-33), estimating the teleport distance (34) or number of teleports (35) needed to reach the orb, or even estimating a metric distance in metres (36).

(32) p03: ok not too far. not too far

(33) p24: and then, quite a long distance again

(34) p45: uh, yeah, yeah short hop

(35) p03: er go four times

(36) p42: and forward about 20 or 30 metres

Some descriptions in the corpus were also quite vague, such as generic underspecified locations expressed with demonstratives like *here* and *there*.

(37) p04: erm somewhere here

(38) p10: yes, it's right there

or perspectiveless topological relations (see Levinson 2003) describing the relation of the orb to other objects in the scene, but without any explicit reference point:

(39) p30: I think, it's like sort of stuck to the tree

(40) p26: you're like on top of it

(41) p26: the one right next to you

Demonstratives were also sometimes combined with the Seeker physically pointing with controller to ask for clarification (Figure 4-15). This could help establish common ground, for instance when introducing a field-based system, such as in (43).

(42) p20: here, this one?

(43) p46: is that 10 o'clock?

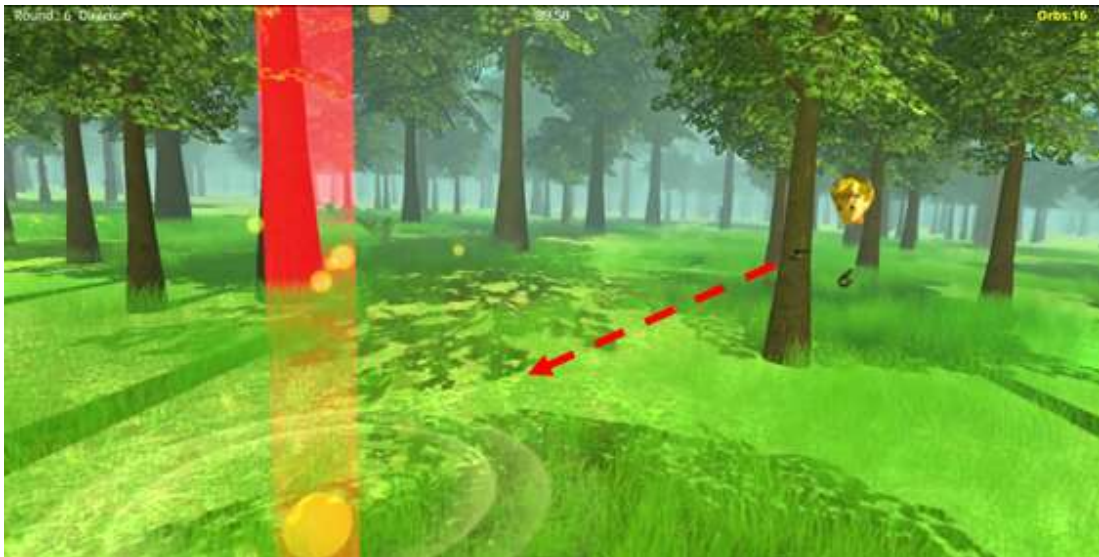


Figure 4-15: Screenshot of the Seeker pointing with his controller to ask for clarification (seen from the Director's perspective).

Similarly, participants would also sometimes use more generic locative descriptions as an attempt to establish common ground, for instance when the Director lost track of the Seeker and they had to find each other again. In such situations relying on topographic landscape features could help players find each other again:

(44) p01: I'm on the top of a mountain

(45) p18: where there are like a bunch of trees

(46) p20: are you in like a large open area?

Such descriptions appear rather unsystematic and are in many ways similar to the ‘figural’ description scheme found in studies using the Maze Game (Garrod & Anderson 1987; Garrod & Doherty 1994; Nölle, Fusaroli, et al. 2020) discussed further above. Concrete features of the environment are used to establish an idea about the location of a player. In the coding scheme, vague locative descriptions like this, which did not rely on a clear coordinate system, are distinguished from geocentric landmark-based descriptions that specify the position of an orb relative to a landmark (e.g., examples 4-8 above). Some descriptions could also be classified as ‘topological relations’ (Levinson 2003: 71–74) that express no angular or coordinate information such as in the following examples:

(47) p30: I think, it's like sort of stuck to the tree

(48) p26: you're like on top of it

(49) p26: the one right next to you

In sum, the OrbHunt game evoked a wide variety of spatial referencing strategies. In addition to describing the location of an orb with a figure-ground relation relying on a specific FoR, players also used field-based strategies, units of distance and other strategies to coordinate with each other. In the following section I will show how these strategies were distributed across both environmental conditions and test whether the environment affected FoR usage.

Impact of environment

Before focusing on differences in FoR, we can consider all spatial referencing strategies discussed above, which are depicted below for both environments:

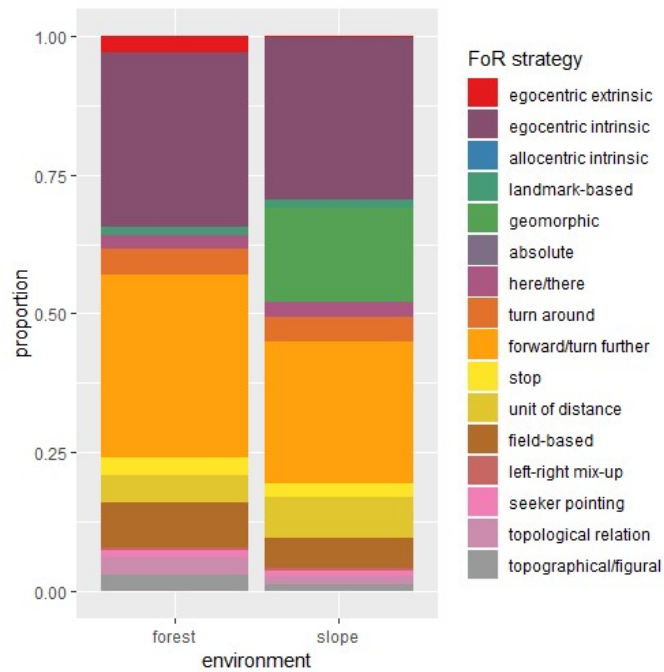


Figure 4-16: Proportions of all coded referencing strategies across both environmental conditions.

Figure 4-16 suggests that dyads in both environmental conditions used a wide variety of strategies to solve the game. Among the different FoR strategies, egocentric intrinsic descriptions (mostly used to express an orb location with respect to the Seeker's left or right) were most frequently produced across both environments. Similarly, variation of *go ahead* or *turn further in the same direction* appear to have been central to solving the game and show a similar frequency. From this graph, it is already apparent that the most striking difference between the two environments is in the usage of the geomorphic FoR, i.e. geocentric descriptions relying on the *up*, *down* and *across* axes in the slope environment, which seems to be in line with the prediction that geocentric affordances could affect FoR usage.

To test whether the environment had a significant impact FoR use, we can use a simpler FoR typology (see Table 3-1 in chapter 3) that allows us to treat FoR strategy as a binomial outcome variable with levels 'egocentric' and 'allocentric', as depicted in Figure 4-17 below.

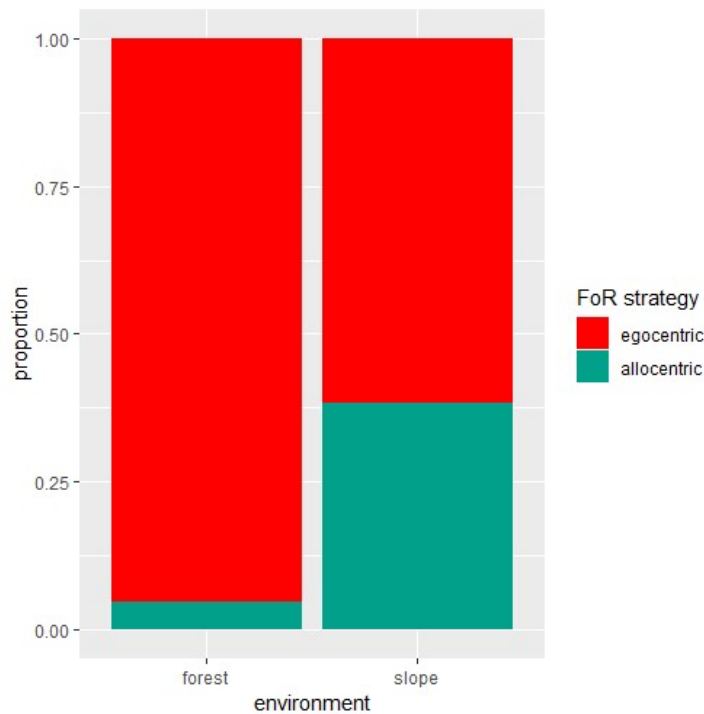


Figure 4-17: Proportions of FoR strategies across both environmental conditions.

A binomial generalised mixed effects model was constructed in R (R Core Team 2019) with the `lme4` package (Bates, Mächler, et al. 2015), with FoR strategy predicted by an interaction between environment and round. The model included random intercepts for individual participants nested in pairs after a model with the maximal random effects structure did not converge.

The regression model (described in Table 4-2) yields a main effect for environment ($\beta = 2.261, p < 0.001$), indicating that dyads playing the slope environment showed a lower propensity to use the egocentric FoR compared to dyads playing the forest environment, which in turn were more likely to produce allocentric spatial descriptions on the slope. Round did not affect the choice of FoR strategy ($p = 0.117$), and there was no interaction between environment and round ($p = 0.193$), suggesting that these differences emerged rapidly in the first round.

Table 4-2: Fixed effects of the regression model used to analyse FoR strategies with respect to environment and round.

	Estimate	SE	p	
Intercept	-1.843	0.242	< 0.001	***
environment (= forest)	2.261	0.479	< 0.001	***
round	-0.044	0.028	0.117	
environment \times round	0.073	0.056	0.193	

4.2.4 Discussion

The results from experiment 1 suggest that environmental affordances did affect the usage of FoR strategies in the OrbHunt game. Comparing dyads' performance between environments indicates that the task was highly similar in both environments. The most popular FoR strategy was egocentric, as was expected for English-speaking participants. However, even though egocentric strategies were perfectly viable on the mountain slope, dyads produced significantly less egocentric and more allocentric descriptions when playing this environment.

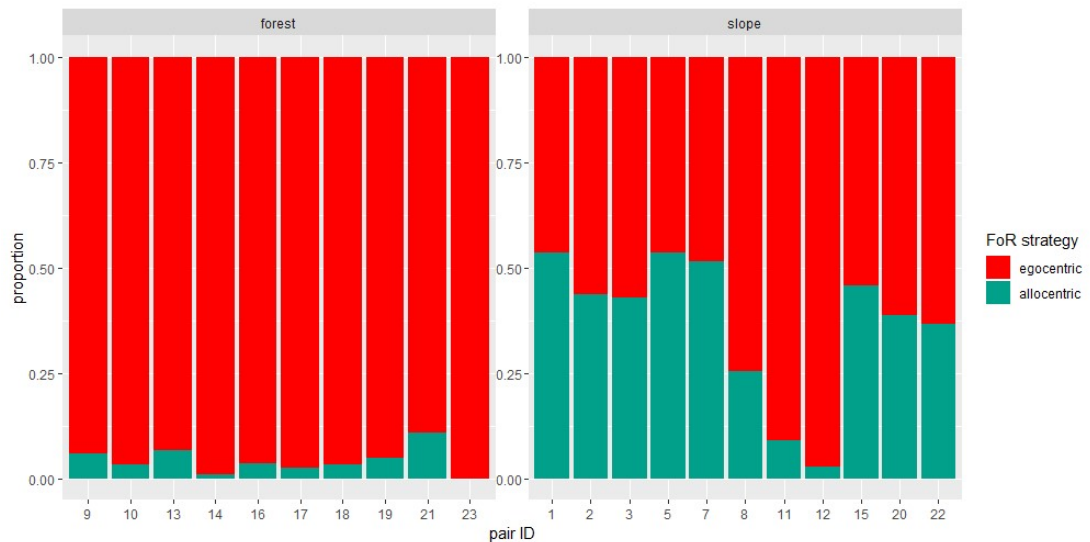


Figure 4-18: Proportions of FoR strategies for individual pairs across both environments.

If we look at the strategies for individual pairs (Figure 4-18), we can see that some dyads on the slope (e.g., pair 11 and pair 12) look nearly identical to pairs playing the forest condition. However, many pairs playing the slope combined egocentric descriptions with allocentric descriptions. More specifically, they produced geomorphic spatial relations that made use of the salient up-down axis of the hill to communicate orb locations.

If an egocentric strategy is viable, why did dyads playing the slope make use of geocentric strategies? One reason could be that geocentric strategies are more reliable. The dialogues contained a total of 35 instances where participants miscommunicated or mixed up *left* and *right*, which corresponds to about 2% of egocentric intrinsic descriptions:

(50) p09: ok, turn 90 degrees to the right

p09: ah, ok, the other way

(51) p10: huhuh

p09: I'm not sure what's happened, cause I think-

p10: oh, no, no, I made a mistake, I was left

In these instances, the Director tried taking the perspective of the Seeker, but mixed up the egocentric directions in which the Seeker should turn. This could for example happen if the orb was to the right-hand side of the Director, but to the left-hand side of the Seeker. This particular pair even suspected that there was a manipulation in the design of the game, because of how frequently they failed to coordinate properly:

(52) p09: I think now it's been reversed, since when I say left you turn right, I think

the directions have been reversed

p10: huhuh, ok, here

p09: ok, so face downhill, a little more

Subsequently, they switched to a geocentric strategy (*downhill* in (52)) that seemed to work more reliably for them. While this example is of course anecdotal, it illustrates the advantage of environment-based FoR, namely that they are viewpoint-independent, and can thus be easily applied as long the environmental anchors needed to construct spatial relations (e.g., the *downhill* direction) are known to both interlocutors. This might explain why many dyads on the slope found it useful to utilize geocentric descriptions even though egocentric descriptions would have been perfectly viable to solve the task.

Another finding of experiment 1 was that the tendency to use geocentric descriptions on the slope emerged rapidly in the first round. Participants started to use *uphill-downhill* descriptions beginning from the first round as soon as they encountered the slope and the proportion of egocentric to allocentric descriptions remained more or less constant over time.⁷ Since English allows for both strategies, this suggests that topographic affordances are enough to trigger the geocentric FoR if it is available to speakers of a language. This is consistent with findings from different field-sites suggesting that speakers of languages that are otherwise associated with the use of an egocentric FoR strategy, will tend to use a geocentric strategy when they frequently move in or engage with specific environments (e.g., Shapero 2017; Calderón, De Pascale & Adamou 2019).

⁷ It should also be noted that the environments contained no 'surprises'. From the first round, participants encountered the full range of potential referents, since all spawn areas in both environments allowed to quickly gather an impression of each map (each spawn location was located on the same slope or surrounded by several trees in the forest). Participants could therefore quickly decide from early on how they wanted to refer to locations in the game and these descriptions would mostly remain reliable throughout the experiment.

If left-right can be easily confusing, why is it so widespread and dominant among many languages spoken in globalized, post-industrial societies and why does it seem to culturally override geocentric systems that appear easier to learn and use (Haun et al. 2006; Bohnermeyer et al. 2015; Shusterman & Li 2016)? One hypothesis could be that flexibility of these FoR strategies plays a role: An egocentric left-right strategy can be applied across many different environments. As discussed earlier, some dyads playing the slope environment produced very similar descriptions to dyads playing the forest environment. However, the reverse is not possible. Geocentric FoR strategies depend heavily on the presence of certain affordances in the environment, such as salient topographic features. Thus, they cannot be easily applied to different environments. As humans became more mobile and started to inhabit more complex environments this could have given an advantage to egocentric strategies, which would spread particularly well in languages spoken across larger areas and many different environments. Experiment 2 was designed to address this flexibility hypothesis.

4.3 Experiment 2: FoR flexibility when switching environments

4.3.1 Outline and prediction

Experiment 2 was designed to investigate how flexible the strategies that participants used in experiment 1 would be across different environments. To this end, the OrbHunt game was modified in such a way that participants would switch environments in the middle of the game. The prediction was that dyads playing in an environment that would afford the egocentric FoR, like the forest, would keep using a predominantly egocentric strategy even if they would move to an environment that topographically afforded geocentric descriptions, like the slope. By contrast, dyads starting out in on the slope will have to switch from a geocentric to an egocentric strategy if they switch to the forest that does not feature salient topographic features. This would show that the egocentric strategy is more flexible and could be used across different environments, while the geocentric FoR highly depends on the environment in respect to which it has been grounded.

4.3.2 Methods

Participants

40 native-level English speakers ($N= 20$ dyads) were recruited via Edinburgh University's MyCareerHub service and received £5 in compensation for taking part in the study. Like for experiment 1, participants were screened to make sure that they would be comfortable with the VR setup.

Materials

Participants were presented with the same OrbHunt game as in experiment 1. The only difference was that participants switched environments mid-way through the experiment. In the 'forest-slope' condition, dyads played a first block of five rounds in the forest environment before switching to the slope environment for another block of five rounds. By contrast, dyads in the 'slope-forest' condition would play the first block on the slope before being transferred to the forest for the second block. Dyads were randomly allocated to one of the two conditions with 10 dyads in each. Instructions remained the same, except that participants were informed that they would switch environment after five rounds. They were however not told what the environments would look like.

Within the game, once the time of the fifth round ran out, a prompt would appear in both players' headsets informing them that they would now switch environments. Once both players confirmed with the trigger on their controller, they would be transported into the second environment, where they had to confirm once more to begin the sixth round (Figure 4-19).

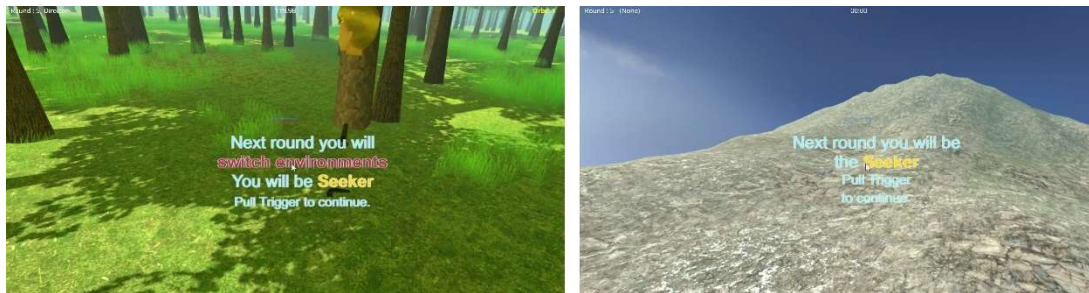


Figure 4-19: Screenshots showing a player switching environments in the forest-slope condition. After round 5 in the first environment has finished, a floating message appears to inform each player about the upcoming environment switch (left image). Once both players have confirmed, they are transferred to the second environment for rounds 6-10 (right image).

Procedure

Besides the addition of the two conditions, the procedure was the same as for experiment 1 (see section 4.2.2.3). Participants entered the lab and provided written consent and received instructions on how to play the OrbHunt game. After familiarising themselves with the mechanics in an empty training environment, players would then proceed to play 5 rounds of the OrbHunt game in each environment for a total of 10 rounds. The block order in which they encountered the two environments depended on the condition they were randomly assigned to.

4.3.3 Results

The recorded dialogues for experiment 2 were coded with respect to the same spatial referencing categories as the dialogues for experiment 1 (see Sections 4.2.3.2 and 4.2.3.3). All coded dialogues are available under https://osf.io/4azur/?view_only=3d977abf51964087a1aa61c2f28dd233. Before looking at the difference between the two conditions, we can consider how the data from experiment 1 compares to experiment 2:

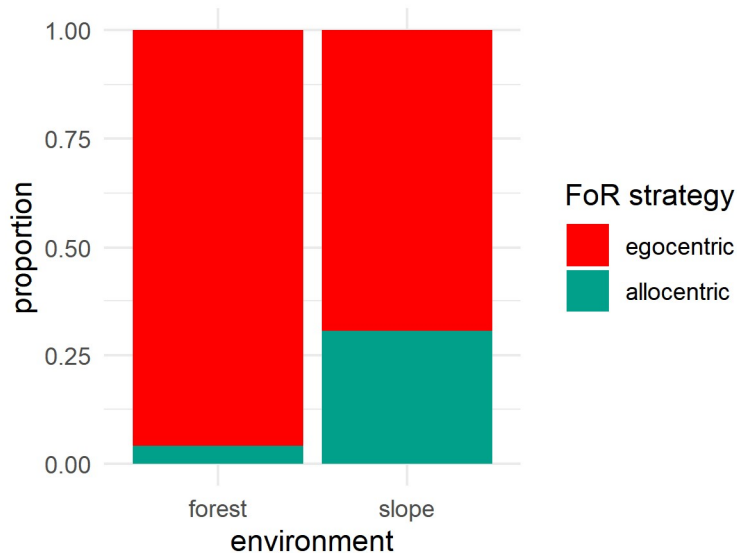


Figure 4-20: Proportions of FoR strategies across both environments.

The distribution of FoR strategies across the two environments appears very similar to experiment 1 (Figure 4-20). However, in contrast to experiment 1, participants switched environments in the second block of the game, which led to shifts in strategies after the fifth round. Considering the impact of condition (block order), according to our prediction, the distribution of FoR strategies should appear flat across both blocks for the forest-slope condition and resemble a step function for the slope-forest condition. This is because we expected dyads to retain their egocentric strategy across both blocks in the forest-slope condition, while dyads in the slope-forest condition were expected to switch from a more geocentric to an egocentric strategy in the second block. Inspecting the results (Figure 4-21), we can observe that in the slope-forest condition, dyads seem to indeed switch strategies, showing a clear step after the first block. However, the same is true for the forest-slope condition, which shows essentially the inverted picture. Dyads starting out in the forest used a larger percentage of egocentric descriptions in the first block before switching to a more allocentric strategy in the second block, contrary to our predictions.

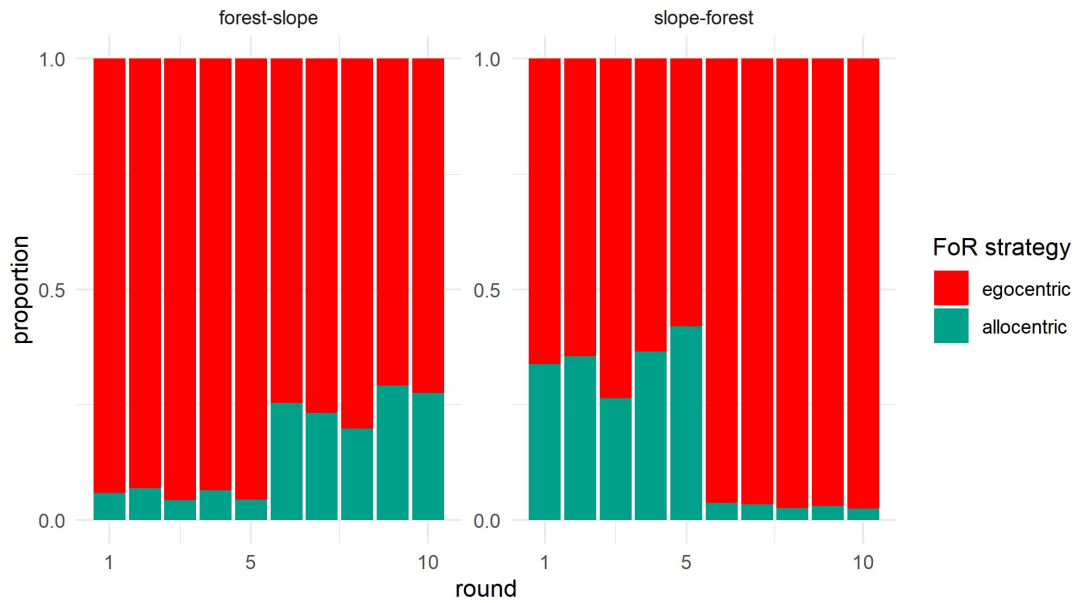


Figure 4-21: Proportions of FoR strategies over time (round) for both conditions/block orders.

Next, the same binomial generalised mixed effects model was fit to the data to test whether the impact of environment reflected experiment 1. The model included FoR strategy predicted by an interaction between environment and round and random intercepts for individual participants nested in pairs (described in Table 4-3). The main effect for environment was replicated ($\beta = 1.219, p < 0.01$), although the effect is slightly smaller than in experiment 1. For experiment 2, the model further yields a significant interaction between environment and round ($\beta = 0.245, p < 0.05$) and a marginal effect for round alone ($\beta = -0.059, p = 0.07$). This suggests that independently of the conditions, participants used similar strategies as in experiment 1, which differed in response to the contrasting environments.

Table 4-3: Fixed effects of the regression model used to analyse FoR strategies with respect to environment and round.

	Estimate	SE	p	
Intercept	-2.253	0.415	< 0.001	***
environment (= forest)	1.219	0.470	< 0.01	**
round	-0.059	0.033	0.071	.
environment \times round	0.245	0.099	0.013	*

To test whether condition, i.e. the order in which environments were presented across both blocks, affected the choice FoR strategy, the `lme4` package (Bates, Mächler, et al. 2015) was used in R (R Core Team 2019) to fit a binomial generalised mixed effects model to the data. The model takes FoR strategy as binomial outcome variable predicted by an interaction between environment and block and accounting for random

slopes for block.⁸ With this we can test whether block order affected participants' FoR choices and whether predominantly egocentric strategies that emerged in the forest were flexible enough to be retained after switching to the slope-environment. The model specification is given in the following formula:

$$FoR\ strategy \sim condition * block + (1 + block | pair\ ID)$$

As would be expected after inspecting Figure 4-21, the model (described in Table 4-4) only yields a significant interaction between condition and block ($\beta = -4.627, p < 0.001$). This simply reflects that the order in which participants encountered the environments affected their strategy, i.e. if they started out in the forest, they would use significantly more egocentric descriptions in the first block and switch to a more allocentric strategy in the second block, and vice versa for the slope-forest condition. However, this switch happened inversely to the same extent in both conditions, suggesting that the simple effect of environment on strategy from experiment 1 was not modulated by switching environments. Instead participants in both conditions readily adapted their referential strategies to the environment at hand.

Table 4-4: Fixed effects of the regression model used to analyse FoR strategies with respect to condition (block order) and block.

	Estimate	SE	<i>p</i>	
Intercept	-2.520	0.341	< 0.001	***
condition (= forest-slope)	-0.034	0.675	0.960	
block (= 1 st block)	-0.131	0.362	0.718	
condition × block	-4.627	0.659	< 0.001	***

4.3.4 Discussion

Experiment 2 tested whether the order in which participant pairs would switch between environments would affect their choice of FoR strategy in the OrbHunt task. Such a difference between conditions was not found. Most dyads readily shifted from an egocentric to an allocentric/geocentric FoR strategy when switching from the forest to the slope environment. Thus, experiment 2 could not provide evidence for a flexibility advantage of an egocentric left-right strategy over an allocentric uphill-downhill strategy.

A possible explanation is that experiment 2 did not actually test the conventionalization of an emerging FoR strategy as such. Instead, participants could readily rely on descriptions that their native language

⁸ A more complex model including an interaction term for block and round per block was also constructed using the following formula:

$$FoR\ strategy \sim condition * block * round_{block} + (1 + block * round_{block} | pair\ ID)$$

However, this model did not converge and there was no indication that round per block did drive any effect (see also Figure 4-21 showing little changes within blocks).

already supported. In other words, there was no cost associated with switching between strategies, since it could be assumed that one's partner would understand both egocentric and geocentric descriptions from the get-go given that English supports both variants. If strategies were to be grounded in interaction first, there would be a cost to shifting to a different strategy upon switching environments. In such a situation, a more flexible strategy that does not depend on environmental cues would have an advantage, since it could simply be retained at no additional cost.

I therefore decided to adapt the task in a way that allowed simulating the emergence and conventionalization of spatial referencing systems that were not previously available to the players. In the following chapter, I will present a new version of the OrbHunt game, in which participants must ground a novel communication system from scratch instead of using their native language to describe locations in the game. Such an 'experimental semiotics' approach (Galantucci & Garrod 2011) can be used to simulate the emergence of communication systems in the lab – in this case modelling the origin of spatial referencing systems that rely on different FoR strategies. This allowed me to test whether the environment affects the evolution of spatial referencing systems more generally as well as the impact of switching environments on different FoR strategies when these must first be grounded and conventionalized.

5 Emergence of a novel spatial communication system in the lab

5.1 Introduction to experiment 3

Experiment 2 presented in chapter 4 suggests that native speakers of a language, which supports the usage of both the egocentric and geocentric FoR, will readily switch between these reference frames across VR environments with differing environmental affordances. Since speakers were already familiar with their L1's conventionalized solution to conceptualise spatial relations in both ways, there was very little cost associated with switching strategy with the same partner when moving to a different environment. Experiment 3 presented in this chapter, therefore takes another approach, simulating the emergence of artificial spatial referencing systems. This approach is usually referred to as Experimental Semiotics (Galantucci, Garrod & Roberts 2012) and has been one of the key methodological innovations in language evolution research discussed in [Paper 1](#) and also used in the Systematicity study presented in [Paper 2](#). The idea is to provide dyads with a collaborative problem that they can only solve via communication while not allowing them to use their native language. Instead, they must rely on a novel communication channel, which is usually graphical or acoustic (Galantucci 2009; Tamariz 2017). For the OrbHunt game, I decided to use a graphical medium, namely visual signals that would appear in the game and could be sent by the Director to the Seeker. Participants were instructed to use this medium to communicate with each other. The signals were designed, so that participants had to negotiate what they wanted them to refer to without being given any predefined form-meaning mapping. In other words, in this version of OrbHunt, dyads had to ground symbols representing spatial descriptions required to solve the task and conventionalise them effectively to succeed in the game. While the game was identical to experiment 2 otherwise, this manipulation should make it costly to switch strategies, since every meaning has to be negotiated in interaction. This allows testing whether the carry-over effects that could not be observed in experiment 2, would result from there being a cost to switching strategies due to renegotiation. If a strategy relying on the egocentric FoR is indeed more flexible, then dyads should simply retain this strategy across both environments, while dyads who evolved a geocentric strategy would have to abandon this strategy and negotiate new mappings if they encountered an environment that did not feature the environmental anchors the geocentric system was grounded on (i.e. switching from slope to forest).

Since there is no spoken dialogue in this version of the OrbHunt game, one challenge lies in interpreting the meaning of signals produced by the participants. In this chapter, I will first present the design

of the task, specifically the novel communication mechanism. I will then show how a mixed-methods approach relying on data from a questionnaire and geometric movement data taken from the game engine can be used to infer underlying FoR strategies.

The results of experiment 3 reveal that it is indeed possible to study the emergence of spatial communication systems in the lab, as dyads successfully managed to establish form-meaning mappings that allowed them to score in the game. Furthermore, a strong bias to establish signs relative to the egocentric looking direction of the Seeker emerged in both conditions. I will discuss the possible origins of this bias as well as some problems with the design of the OrbHunt task itself that are important to take into account when trying to model the evolution and variation of spatial linguistic referencing systems.

5.2 Participants

42 native-level English speakers ($n = 21$ dyads) were recruited via Edinburgh University's MyCareerHub service and received £5 in compensation for taking part in the study. None of the participants had participated in experiment 1 or 2. One additional participant pair was unable to complete the task as intended due to technical difficulties and was therefore excluded from the analysis.

5.3 Materials

Like experiment 2, half of the dyads were assigned to one of two conditions. 11 dyads played the slope-forest condition, and 10 dyads played the forest-slope condition, since one pair was excluded. They played five rounds of the OrbHunt game before switching environments for another five rounds. Each round again lasted for 120 seconds and the experiment took approximately 30 minutes in total. The crucial difference to experiment 1 and 2 was that participants were not allowed to use any language or other sounds to communicate in the game. Instead the Director was able to send pictorial signals which appeared on both players' in-game visual display. During each round, the Director could use either of the two grip buttons on their controller (the one that was not used for displaying time, score and role) to rotate through 8 available signs on click and send them to their partner by pulling the trigger for confirmation. The chosen sign would then appear in the Seeker's headset, approximately in the centre of their visual field and would remain visible for two seconds, before vanishing. Simultaneously, it would also appear in the Director's view for two seconds as a form of feedback, confirming that the Seeker had received the message. A minor change to the previous experiments was that orbs could now be collected automatically upon contact with the controllers collider (i.e. if players 'touched' them), while in the previous versions of OrbHunt they had to be manually 'grabbed' and 'collected' by pressing buttons on the controller. This

made it easier for participants to remember the multiple button assignments, and furthermore experiment 1 and 2 revealed that the grab-and-collect mechanic only added an additional unnecessary level of difficulty, since several participants required multiple attempts to understand how to collect an orb.

For an initial pilot, this novel communication medium comprised 8 arbitrary signs that were designed to appear random and unfamiliar, but easily distinguishable via distinct colours and shapes (see Figure 5-1).

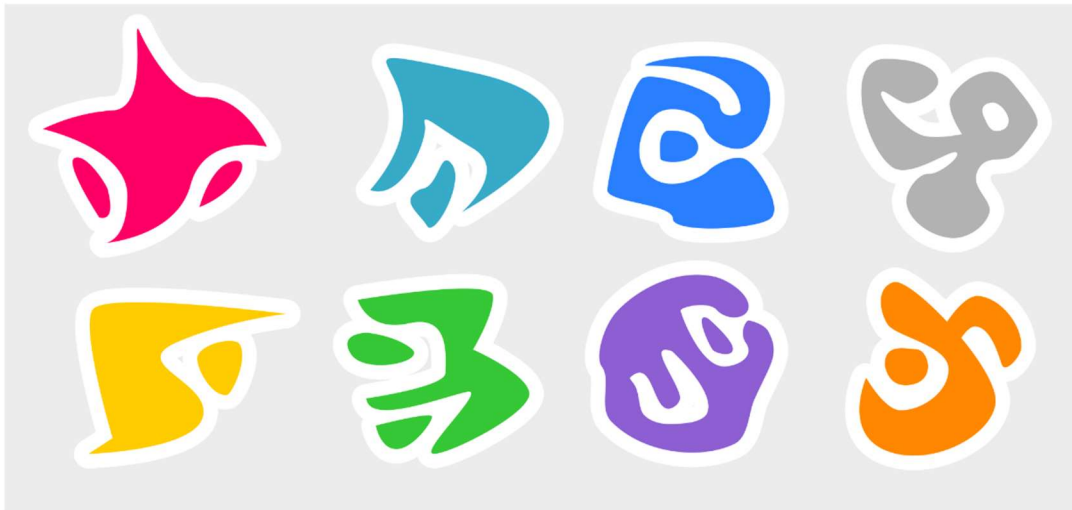


Figure 5-1: Signalling space for the pilot experiment. Each signal had a distinct colour and shape and appeared with a white border in the VR game to set it off clearly against the background.

However, at the pilot stage ($N=18$) it became clear that there was an extremely strong bias for grounding egocentric communication systems, i.e. mapping meanings such as *left*, *right*, or *forward* onto the signs. In the pilot, out of 9 dyads that started on the slope, not a single dyad attempted grounding a geocentric system by mapping such as *uphill* or *downhill* onto the signs (in contrast to experiment 1 and 2, where geocentric descriptions occurred frequently on the slope). This could either be due to a strong cultural bias for egocentrism that became amplified in this task or it could be a problem with the signals themselves. Although none of the signs was intrinsically designed to afford specifically for *left* or *right*, any left-right asymmetry could easily be exploited to motivate such a meaning. For experiment 3, I therefore decided to facilitate the grounding of geocentric signal-meaning mappings by making half of the signs more iconic regarding features of the environment. For instance the turquoise and purple triangular shapes in Figure 5-2b could be used to establish an *up/down* distinction, while the blue and orange sign could be used to refer to landmarks such as trees or the sun, respectively. These iconic properties do of course not preclude grounding an egocentric sense for these signs: The blue and turquoise sign could just as well be used in the

forest, e.g., to communicate *forward*, and the mapping to a meaning does not have to rely on iconicity and can be purely arbitrary.

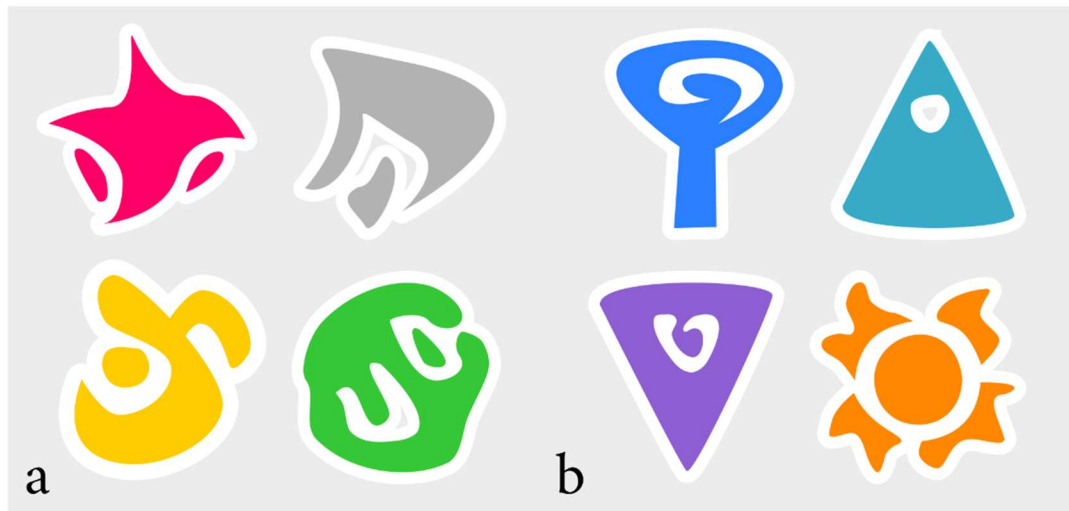


Figure 5-2: Signalling space used for experiment 3. Half of the signals appear arbitrary (a), while the other half feature iconic shapes that could be used to ground mappings to features of the environment (b).

To not make any sign inherently more attractive than another and exclude order effects, the list of signs appearing on the Director’s controller would be initialized at a random position each time they opened it. This effectively also limited the frequency at which signs could be sent, since the Director usually had to rotate through several signs before reaching the sign they intended to send to the seeker. This limitation was also established to discourage players from repeatedly sending the same signal in rapid succession, which would be more akin to ‘steering’ the Seeker in real-time (e.g., “go forward, left, forward, forward, stop!”) rather than producing location descriptions.

The instructions for the game were the same as for experiment 2, except for the fact that participants were instructed to not talk during the experiment. They were told that instead they could only use a set of signs that did not have any inherent meaning to communicate orb locations in the game. The experimenter made sure to stress that it was the participants themselves that had to decide what the signs meant and that creating such a novel communication system would involve a lot of trial and error, encouraging them to not give up if they wouldn’t be able to score on the first few trials. (See [Appendix C](#) for instructions)

After completing the OrbHunt game, participants were presented with a questionnaire ([Appendix D](#)). The questionnaire asked participants to describe what the signs they had used to communicate in the game meant. For each sign, they could provide two responses, 1) a short description explaining how they used the sign, and 2) a label capturing the intended meaning in a word or simple phrase. Participants

could also comment on their general strategy for establishing successful communication and what worked/did not work so well when interacting with their partner.

The predictions for experiment 3 were as follows:

1. Like in experiment 1 and 2 participants are expected to differ in their FoR strategies across environments, with less dyads using an egocentric FoR if they start the game in the slope environment.
2. Due to there being a cost to establishing a new FoR convention, we expect the following differences between conditions: Dyads switching from forest to slope should retain their predominantly egocentric strategy, while dyads switching from slope to forest have to re-negotiate their original strategy, since geocentric strategies are less flexible across environments and ineffective in the forest.

5.4 Procedure

Procedure was the same as experiment 2. Participants came into the lab, where they were instructed and then played 10 rounds of the OrbHunt game, switching environments after five rounds (either slope-forest or forest-slope). Experiment 3 differed in that participants were not allowed to talk to each other and the recorded video data contained no dialogue. Instead, after completing the OrbHunt game, each participant completed a short questionnaire to provide qualitative insights into signals' underlying meanings and their underlying referential strategies.

5.5 Results

5.5.1 Performance

Unsurprisingly, participants' inability to rely on language and the challenge of bootstrapping a novel communication system made the task much more difficult. This resulted in lower scores in comparison to experiment 1 and 2 in both conditions, with median scores of 12 and 16 orbs, respectively (Figure 5-3). Only two pairs in the forest-slope condition managed to achieve a score that would be considered average for experiment 1 and 2, while other pairs scored much lower, including several pairs collecting less than 10 orbs within 10 rounds and one pair even failing to solve the task altogether, resulting in a score of zero collected orbs.

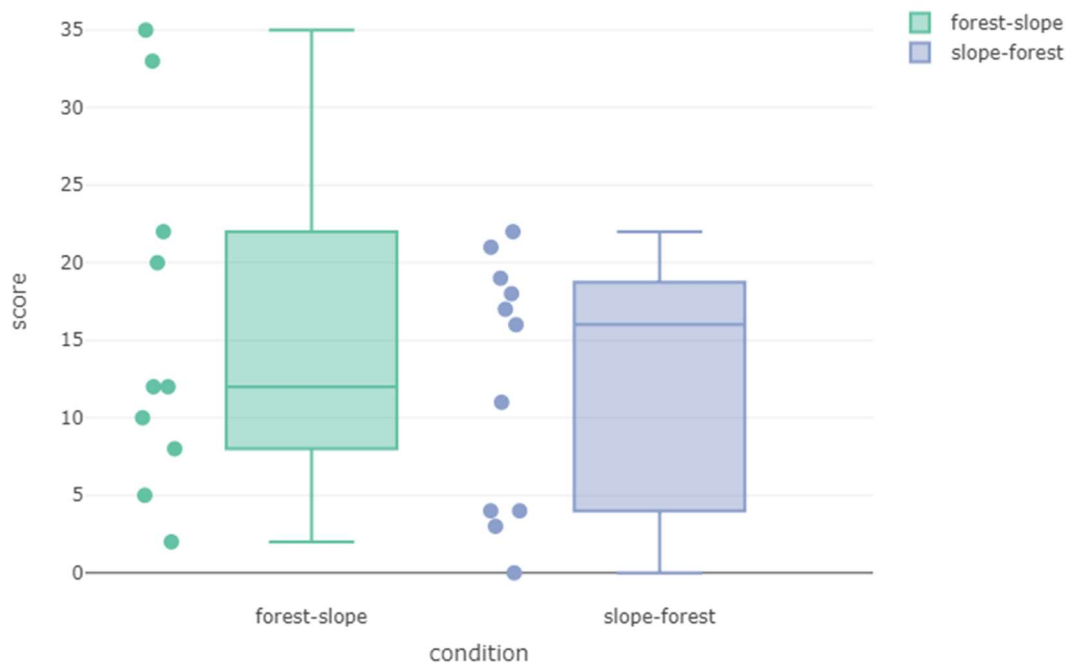


Figure 5-3: Boxplots showing the number of collected orbs for each pair across both conditions. Dots represent scores for individual pairs.

Looking at performance over time, we can observe that many pairs started relatively slowly and often needed several rounds until they successfully collected their first orb (Figure 5-4). By contrast, scores rise at a faster rate in later rounds, suggesting some increase in efficiency at later stages of the game, possibly due to dyads conventionalizing a grounded spatial referencing strategy. In the following sections, I will analyse these strategies using a combination of qualitative and quantitative analysis.

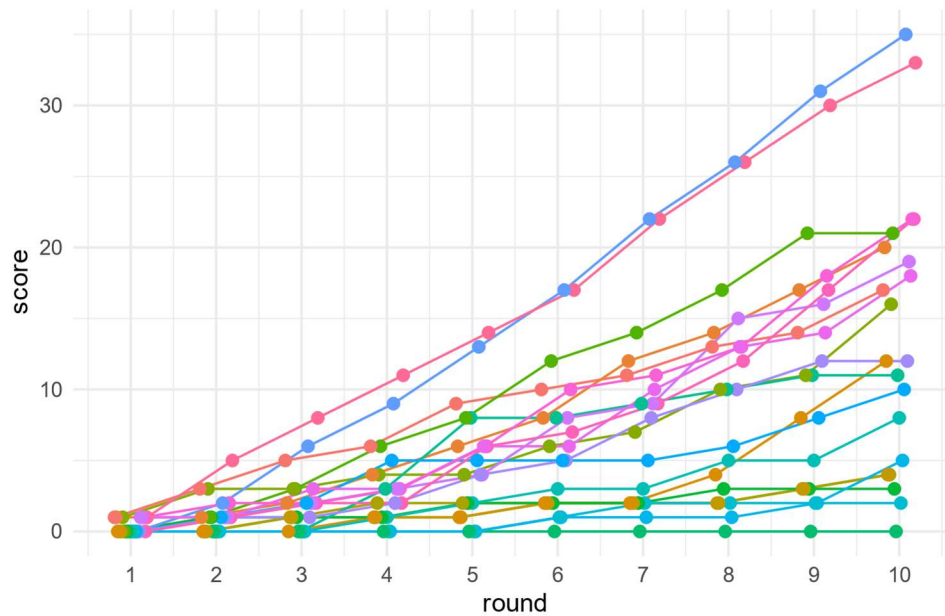


Figure 5-4: Number of collected orbs over time for all 10 rounds of the experiment. Colours represent individual dyads and dots are offset horizontally to make overlapping scores visible.

5.5.2 Qualitative analysis of the communication medium

Solving the symbol grounding problem

As indicated by the scores discussed above, dyads were able to solve the task and communicate via the novel visual communication channel. Reviewing the video footage suggested that many pairs used an ad hoc grounding strategy. For instance, the Director would send a sign and observe where the Seeker would move in response. The Director would then keep using the same sign whenever they wanted to communicate the same direction/location again. Challenges included remembering what a sign was used to express. As one participant remarks in the post-experiment questionnaire: “having the symbols signify as something did not work as there were too many to remember. Hence, instead we would use the same symbol if it was going in the correct direction and a different one if it is wrong” (p30). Another challenge was ambiguity resulting from the same symbol being used or interpreted differently: “Got directions well but used multiple symbols to represent same thing which caused confusion” (p07).

The videos showed that similar to experiment 1 and 2, it became clear to both players early on that a good strategy was for the Seeker to wait for instructions from the Director and then move in response to these:

- (53) “I think a reliable strategy is for the seeker to just stand still until an instruction is sent. Maybe, the seeker could point one of the controller to let the director understand where he\she wants to

go. A reliable strategy for the director is to follow the seeker and always use his\her point of view in order to send good instructions.” (p13)

One way to solve the ‘symbol grounding problem’ (Harnad 1990; Steels 2008) in referential communication via a novel medium is to exploit iconicity in the signalling space (Fay, Arbib & Garrod 2013; Verhoef, Kirby & de Boer 2016). Interestingly, some participants tried to use colour to refer to specific aspects of the environment. Individual participants tried to use the red cross-like sign to communicate *stop*, or *wrong way* and several participants used the orange signal to signify that the orb would be close by (presumably due to the resemblance in shape and colour). However, features of the environment could also be used. For example, participant 11 tried to use the green sign in Figure 5-2 to refer to a patch of grass in the forest environment while participant 09 used the same sign on the slope to refer to the lower area of the mountain that was greener and featured more grass. Similarly, participant 29 used the blue sign to refer to what they describe as “the sea” by which they meant the direction of the area at the foot of the mountain (that simply faded into the blue texture of the sky, so as to not reveal the edge of the explorable map). Only two individual participants picked up on the iconic resemblance of the blue sign to a tree, and attempted to use in the forest environment to direct their partner to a tree or ‘cluster of trees’, respectively. However, in none of these cases did their partner share this interpretation in the questionnaire, suggesting that establishing colour or shape-based symbols was not a very successful strategy and rather constitutes failed ad hoc attempts. Only few signs were commonly associated with a specific meaning. I will discuss these in the following subsection, focusing on the underlying FoR strategies.

Frames of Reference

What about the iconic affordances that were by design included in the signalling space? Given the results from experiment 1 and 2, one should expect participants in the slope environment to pick up on environmental properties and at least to some extent use the equivalent to terms such as *uphill*, *downhill* or *towards the sun* to communicate about orb locations.

However, responses from the questionnaire suggest that this did not work as well as expected. Individual participants did attempt to map such geocentric meanings onto signals, but would often give up, since their partner did not understand them or moved in a different direction than intended by the Director. Only a single pair (pair 11) responded that they used the turquoise triangular signal that was designed to be iconic with regards to the uphill direction to actually communicate “going up in the mountain environment” (p22) as well as the purple triangle to communicate the opposite direction “go down” (p21). The pair remark that once they switched to the forest environment, it became “more complicated” (p22), and

they switched to using these two symbols in an egocentric sense, where they would mean “maintain direction” or “switch direction” (i.e. turn around). This is the only clear example of a geocentric strategy that was explicitly abandoned after switching to the forest. The main reason for this being that most participants used these two signs in the egocentric sense expressing *forward* and *backward* from the beginning. A striking 32 of the 42 individual participants included in the analysis reported that they used the turquoise sign to express some form of *forward* or *ahead*, based on the Seeker’s egocentric looking direction, and 30 participants described some kind of meaning associated with *backward* or *turn around* associated with the purple triangle. In their general comments many participants describe using the Seeker’s looking direction as an anchor for communicating directions as a very successful strategy. In other words, ironically, the signals designed to resemble up and down turned out being exapted for egocentric strategies anyway. One participant even explicitly states “for me the symbols were rather confusing, probably arrows would have made the task easier” (p20), suggesting that egocentric ‘arrows’ pointing into a relative direction would be a useful strategy to solve the task, which could help explain why these egocentric mappings were so popular.

Most pairs also grounded some variant of *left* and *right* using a wide variety of signals to do so. The most frequent one associated with these terms was the grey signal, presumably, because it featured the most salient left-right asymmetry. Out of 22 participants, 16 used it to express *right* and 8 used it to express *left*, which could be combined with another symbol signifying the opposite egocentric direction.

Interestingly, a geocentric symbol that appeared more often than *up* or *down* was *sunwards* mapped onto the orange signal, which at least 7 participants report to have used. This might be surprising given that in the previous spoken experiments, the up-down axis were much more frequently used than the crossing sun/moon axis, but it could simply be an artifact of the triangular signals already being reserved for the more salient forward-backward interpretation, making the iconic orange signal the most salient form that remained to map a geocentric meaning onto. In the cases where the sun was used, participants seemed to find the strategy very useful. E.g., participant 11 describes that “We were most successful when we could reference actual objects in the world, like the sun. Forwards and backwards also seemed to roughly work”. Pair 11, the single dyad that properly conventionalised signals referring to the uphill-downhill axis also utilised the sun, and one of the players (p22) remarks that it was difficult to transfer this meaning to the forest environment. In the video recording of their game, one can actually see the Director use the ‘sun’ signal in the forest, after which the Seeker turns around confusedly, since no visible sun can be seen in the sky. After this they stop using the signal.

This provides some indication for geocentric strategies being adaptive to specific environments, but less flexible, resulting in a need to renegotiate the meaning of symbols or ground a completely new (egocentric) strategy. However, this evidence is merely anecdotal, since geocentric strategies were rare, and unlike the previous experiments not frequently adopted on the mountain slope, which made it hard to see how geocentric strategies would fare in the forest environment.

Taken together, the questionnaire confirmed that the scores were not simply the result of luck. In many cases both members of dyad agreed on what the symbols they grounded signified. Participants were able to bootstrap communication systems and create limited artificial spatial referencing systems in the lab. While there were some attempts to motivate symbols based on iconic resemblances related to signals' shape or colour, for many signals, the associated form-meaning associations varied widely across dyads, with some notable exceptions: Two thirds of all dyads used an egocentric strategy, mapping *forward* and *backward* onto two triangular signals, suggesting that this strategy was very successful for solving the task. Similarly, the responses suggest that it was very hard to successfully establish geocentric form-meaning mappings, suggesting that this novel task provided a strong pressure for the egocentric FoR, at least for a population of English speakers. This result parallels an earlier pilot in which even more neutral stimuli (no arrow-like shapes) were used, and egocentric systems emerged in the vast majority.

While the questionnaire data provides a qualitative picture of participants' strategies and the various ways in which form-meaning mappings could be grounded, there are also some problems with this type of data: Several participants did not understand correctly that 'labelling' the sign meant providing its meaning and they instead entered labels such as "star" or "grey pointy thing". Similarly, it cannot be assumed that participants remembered all mappings they used throughout the game. There could be a recency effect or bias to report mostly the signs that were routinized and that a dyad kept using until the end of the game. Lastly, some descriptions were hard to understand, such as when participants describe meanings as 'East' or 'North', but what they actually mean is 'right' and 'forward' based on the Seeker'. In order to run statistics and test predictions 1 and 2, I therefore developed a quantitative approach to infer the underlying FoR strategies directly from the movement data. If these are consistent with the observations from the qualitative questionnaire, we can reasonably conclude that egocentric strategies were indeed more common and successful across all environments and conditions in this version of the OrbHunt game.

5.5.3 Frames of Reference inferred from movement data

While the questionnaire gives a qualitative idea about the strategies participants' developed, it is not the most accurate measure for determining underlying FoRs, given that individual participants could not

necessarily accurately remember how they used each sign during the game and/or how meanings changed over time.

A method to quantify the FoR strategies used to solve the task is inferring the underlying coordinate systems based on actual movement data from the game. The OrbHunt game records each player's position in 3D space as well as their looking direction (the three-dimensional forward vector of the headset's camera in the VR) at any given point in. For each communication event, that is each time the Director sent a signal to the Seeker, we thus know \mathbf{v}_{look} , the current looking direction of the Seeker (i.e. where the Seeker was looking when the director sent the signal), their position, and the direction in which they moved subsequently. The Seeker's movement vector \mathbf{v}_{move} corresponds to the difference vector \overrightarrow{AB} between their position A at time t (the moment they received the signal) and position B at $t + 1$ (the moment after the next teleport). We can then calculate the angle θ between \mathbf{v}_{move} and \mathbf{v}_{look} in order to quantify how the Seeker turned after receiving the signal⁹. The dot product between these two three-dimensional vectors can be defined as:

$$\mathbf{v}_{move} \cdot \mathbf{v}_{look} = \|\mathbf{v}_{move}\| \|\mathbf{v}_{look}\| \cos \theta \quad \text{Eq. 1}$$

where $\|\mathbf{v}\|$ is the Euclidean norm (or L_2 norm) of a vector \mathbf{v} . We can then solve for θ , which lies on the closed interval between 0 and π :

$$\theta = \cos^{-1} \left(\frac{\mathbf{v}_{move} \cdot \mathbf{v}_{look}}{\|\mathbf{v}_{move}\| \|\mathbf{v}_{look}\|} \right) \quad \text{Eq. 2}$$

In order to determine the direction of rotation (i.e. whether the Seeker turned clockwise or counterclockwise), we make θ a signed angle. For θ , a vector \mathbf{a} seen from unit vector \mathbf{n} must rotate counterclockwise to reach another vector \mathbf{b} when:

$$\cos \theta = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|}, \quad \text{and} \quad \sin \theta = \frac{(\mathbf{n} \times \mathbf{b}) \cdot \mathbf{a}}{\|\mathbf{n} \times \mathbf{b}\| \|\mathbf{a}\|}. \quad \text{Eq. 3}$$

We can then set the range to $0 \leq \theta < 2\pi$:

⁹ Note that for the analysis presented here, the movement vector \mathbf{v}_{move} only considers players' horizontal movement along the x-axis and z-axis in the game (while vertical y = 0).

$$\begin{cases} \theta = \arccos \frac{\mathbf{a} \cdot \mathbf{b}}{ab}, & \text{if } \sin \theta \geq 0; \\ \theta = 2\pi - \arccos \frac{\mathbf{a} \cdot \mathbf{b}}{ab}, & \text{if } \sin \theta < 0; \end{cases} \quad \text{Eq. 4}$$

With this method we can set the Seeker's movement in relation to two different coordinate systems, the egocentric FoR, that is the signed angle between the Seeker's movement and their looking direction when the signal was sent, and the geocentric FoR, that is the signed angle between the Seeker's movement and the viewpoint-independent axes of *up/down* and *across*. Within the VR game, these absolute vectors correspond to:

$$\mathbf{v}_{up} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{v}_{down} = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{v}_{sun} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \mathbf{v}_{moon} = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}. \quad \text{Eq. 5}$$

In the slope environment, these directions more or less approximate the axes of the traversable mountain-side and the directions of the sun and moon, while in the flat forest environment, they simply represent four absolute cardinal directions (similar to North, South, East, and West in the real world).

For each sign that a dyad used, we can thus determine, which Frame of Reference is more consistent with their movement pattern by looking at the distribution of angles associated with each coordinate system. For instance, a Director could use the yellow sign to mean *right*. If both partners coordinate this meaning successfully, θ_{ego} , the angle between the Seeker's looking direction and where they moved after hearing the sign, will consistently lie around 90° with respect to the Seeker's egocentric FoR, while θ_{geo} , the angle between the Seeker's movement and any geocentric direction, will vary significantly as the Seeker rotates with respect to the environment (Figure 5-5).

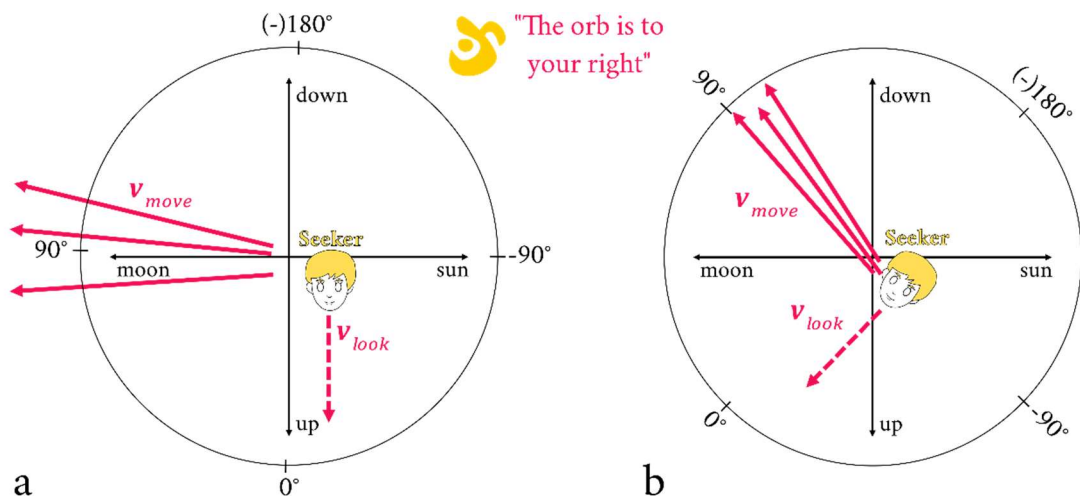


Figure 5-5: Schematic diagram of an egocentric sign expressing the *right* direction based on the Seeker's relative viewpoint. Depending on the Seeker's orientation, their movement vector may coincide with a geocentric direction. In panel a, the sign could also be interpreted as expressing *moonward*. However, as shown in panel b, the Seeker's head orientation can vary when they receive the same signal. Thus, while their orientation changes with respect to geocentric coordinates, their movements remain consistently to her relative right.

By contrast, if a Director maps a geocentric meaning onto a sign, the spread of θ_{geo} will be smaller than that of θ_{ego} . This is due to the fact that the Seeker's looking vector will constantly change as they move through the environment, while geocentric directions remain constant (Figure 5-6).

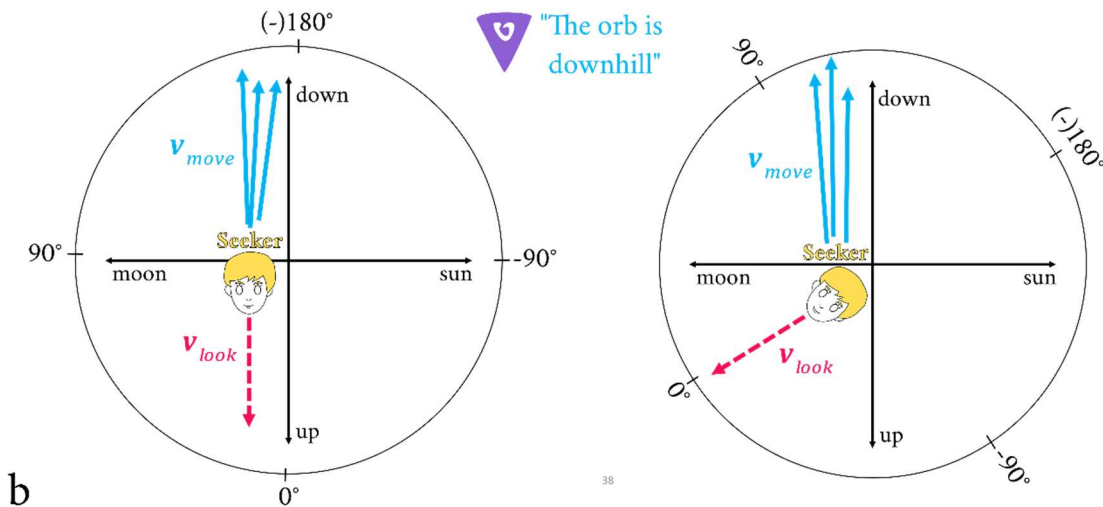


Figure 5-6: Schematic diagram of a geocentric sign expressing *downhill* with respect to topographic properties of the slope environment. In panel a, the sign could also be interpreted as meaning *backward* or *turn around*, given that the Seeker's movements are around 180° clockwise or counterclockwise with respect to their looking direction. However, if the looking direction is not consistently aligned with absolute space, the Seeker's movement patterns will be less consistent with respect to their relative viewpoint. In panel b, movements are closer to a

movement to the *right* rather than *backwards*, while remaining consistent with the *downhill* direction.

Since we are dealing with circular data (where 180° and -180° in the range of $[-180, 180]$ are identical), we have to rely on circular sample statistics (Fisher 1995). For any given sign, per dyad and per environment, we can determine the circular mean $\bar{\theta}$ by representing each angle as a two-dimensional vector composed of its sine and cosine. The circular mean is the mean direction of these vectors, i.e. the vector that results from connecting the resulting angle vectors head-to-toe (Figure 5-7).

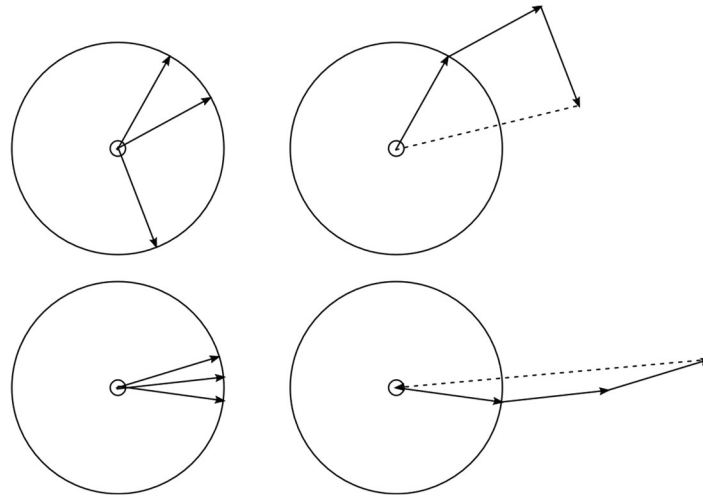


Figure 5-7: Each row shows a different set of circular data represented by solid-line vectors (left) and their circular statistics (right). The direction of the dotted vector is the dataset's circular mean and the length of the dotted vector is its mean resultant length. A larger spread of the data (top left) thus results in a shorter mean resultant length (top right). Figure adapted from Cremers and Klugkist (2018: 5).

We can obtain a measure of spread, the mean resultant length \bar{R} , by dividing the length of the circular mean by the number of vectors. The mean resultant length takes a value between 0 and 1, where larger values correspond to datapoints that are more concentrated on the circle. Conversely, the circular variance Vm is defined as $1 - \bar{R}$. Furthermore, we can obtain a circular standard deviation v , where larger values indicate a larger dispersion using the following formula (see Fisher 1995: 32):

$$v = \sqrt{-2 \ln \bar{R}} \tag{Eq. 6}$$

For analysing the FoR strategies employed by dyads in experiment 3, these sample statistics were computed using the `circular` package (Agostinelli & Lund 2017) in R 3.6.1 (R Core Team 2019). This allows us

to compare the self-reported data from the questionnaire with actual behavioural data from the OrbHunt game itself.

Determining Frames of Reference by movement angle spread

In order to determine which FoR was more likely to underlie a given signal, we can, for each pair, compare the mean resultant lengths of all angles associated with this signal in one of the two environments.

Consider for example the movement pattern of pair 11, the only pair, where both players reported having used the orange sign to refer to the direction of the sun in the slope environment. This pair started playing on the slope and then progressed to the forest (slope-forest condition). If we plot the angles of their movements after using the orange sign with respect to the egocentric and geocentric coordinate systems across both environments (Figure 5-8), we can observe that the geocentric angles are much more concentrated on the slope, clustering around $\theta_{geo} = -45^\circ$.¹⁰ The single outlier ($\theta_{geo} = 14.12^\circ$) corresponds to the movement after the sign had been used for the very first time, suggesting the Seeker did not interpret its meaning correctly at this instance. By contrast, the spread of geocentric angles is much higher in the forest environment.

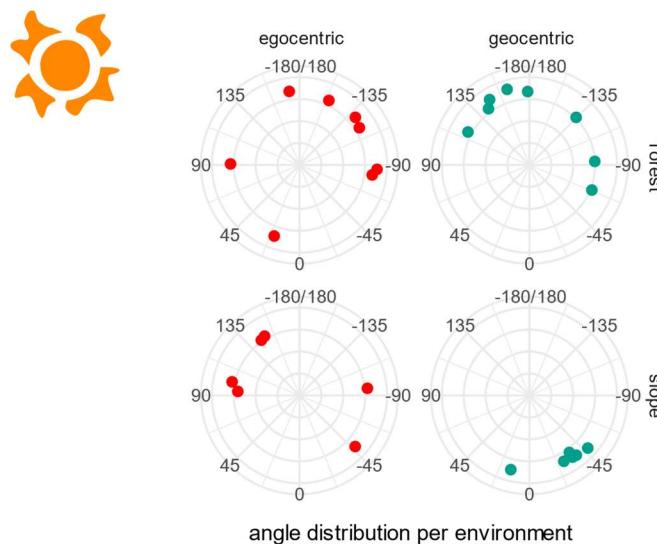


Figure 5-8: Distribution of movement angles associated with the orange sign for a single dyad (pair 11) playing the slope-forest condition.

It may appear surprising that this pair kept using a sign that had been used to communicate *sunward* on the slope after switching to the forest, where no distinct sun was visible in the sky and absolute directions are difficult to keep track of. However, one player reported that they attempted to use it to communicate

¹⁰ Note that these polar plots represent egocentric angles based on the viewing direction if the Seeker was placed in the middle of the circle, like in figures 5-5 and 5-6 above. Thus, forward (0°) is down on the page and right (90°) and left (-90°) appear reversed, as if the reader is ‘facing’ the Seeker.

“go to a sunny area in the forest”, implying a clearing being used as a landmark, while their partner reported that they used it to communicate “left or right depending on partners reaction” in the forest environment. This less consistent way of using the sign in the forest resulted in a larger spread of angles in relation to the geocentric FoR (i.e. the mean resultant length is smaller for movements with respect to geocentric directions compared to movements with respect to the Seeker’s looking direction) . The tables below show the sample statistics for the movement angles after using the orange sign on the slope (Table 5-1) and in the forest (Table 5-2). Note that I provide the angles for both geocentric axes, the uphill-downhill and the across axis separately, since players were moving their headset in three-dimensional space, which could affect their relative orientation to each axis slightly differently (e.g., in Table 5-3 below).

Table 5-1: Circular sample statistics for pair 11 using the orange sign in the slope environment (block 1). $\bar{\theta}$ = circular mean, \bar{R} = mean resultant length, Vm = circular variance, v = circular standard deviation.

Angle relative to	$\bar{\theta}$	\bar{R}	Vm	v
Egocentric viewpoint	136.14	0.32	0.69	86.72
Up-down axis	-118.91	0.94	0.06	19.84
Across axis	-28.91	0.94	0.06	19.84

Table 5-2: Circular sample statistics for pair 11 using the orange sign in the forest environment (block 2).

Angle relative to	$\bar{\theta}$	\bar{R}	Vm	v
Egocentric viewpoint	-127.53°	0.4	0.6	77.45°
Up-down axis	94.19°	0.56	0.44	61.49°
Across axis	-175.81°	0.56	0.44	61.49°

The tables show that in the slope environment, the mean resultant length of the movement angles with respect to the geocentric FoR, $\bar{R}_{geo} = 0.94$, is much larger than the mean resultant length with respect to the egocentric FoR, $\bar{R}_{ego} = 0.32$, i.e. the movement angles are more tightly clustered with respect to geocentric rather than egocentric coordinates. By contrast, in the forest environment, the mean resultant lengths are much more similar as movement angles are more widely spread with respect to both coordinate systems.

Using the same approach, we can look at an example of a clearly egocentric strategy, such as the frequently reported *forward/backward* system relying on the triangular signs depicted in Figure 5-2 above. For instance, both players from pair 20 (slope-forest condition) reported having used the turquoise sign to

express *forward* (i.e. an egocentric intrinsic strategy based on the Seeker's viewpoint). If we plot this pair's movement angles, we can see that the angles cluster around 0° with respect to the Seeker's egocentric FoR, while they are much more spread around the circle with respect to the geocentric FoR (Figure 5-9). This tendency remained constant across both environments, suggesting the pair kept using this strategy as soon as they had established it after some initial trial and error (characterised by egocentric outliers in the slope environment).

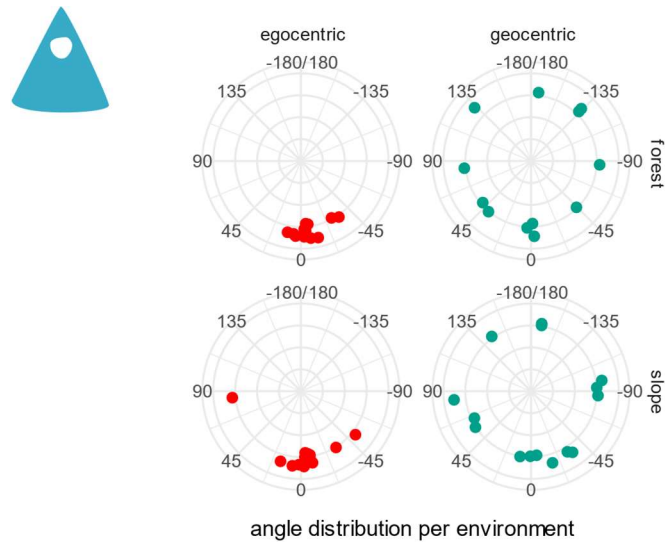


Figure 5-9: Distribution of movement angles associated with the turquoise sign for a single dyad (pair 20) playing the slope-forest condition.

As a consequence of this egocentric strategy, the mean resultant length of the angles relative to the egocentric FoR is much larger in both environments, as can be seen in the tables below:

Table 5-3: Circular sample statistics for pair 20 using the turquoise sign in the slope environment (block 1).

Angle relative to	$\bar{\theta}$	\bar{R}	Vm	v
Egocentric viewpoint	-3.23°	0.90	0.10	26.4°
Up-down axis	-52.82°	0.14	0.86	112.6°
Across axis	-18.40°	0.27	0.73	93.32°

Table 5-4: Circular sample statistics for pair 20 using the turquoise sign in the forest environment (block 2).

Angle relative to	$\bar{\theta}$	\bar{R}	Vm	v
Egocentric viewpoint	-6.75°	0.98	0.02	12.5°
Up-down axis	-92.15°	0.18	0.82	105.94°
Across axis	-2.15°	0.18	0.82	105.94°

Analogously, pair 17 reported using the purple sign to express *backwards/turn around*, i.e. the opposite egocentric direction corresponding to the vector opposite to the Seeker's looking vector. As a result, the movement angles cluster around 180°/-180°, which reflects the Seeker making a U-turn (Figure 5-10). This example shows a more gradual conventionalisation process: While the mean resultant length is larger for the egocentric FoR across both environments, the angles are more spread out in the first block (the forest environment). This suggests that pair 17 went through some trial and error during the first few rounds in the forest environment, but by the time they switched environments they had established a convention that allowed them to reliably communicate the intended direction using the purple sign. Note also, that this example illustrates the sensitivity of \bar{R} with a low number of datapoints: Even though the egocentric FoR appears much more consistent in the forest environment compared to the geocentric FoR, the two outliers in the forward direction around 0° lead to large standard deviation of $v_{ego} = 74.57^\circ$, which is comparable to that relative to the geocentric FoR, $v_{geo} = 77.79^\circ$, which causes $\bar{R}_{ego} = 0.43$ to appear almost identical to $\bar{R}_{geo} = 0.4$. In the second block, there aren't any outliers that stray far from the mean angle, and thus the mean resultant length is much larger for the egocentric FoR ($\bar{R}_{ego} = 0.98$) than for the geocentric FoR ($\bar{R}_{geo} = 0.18$).

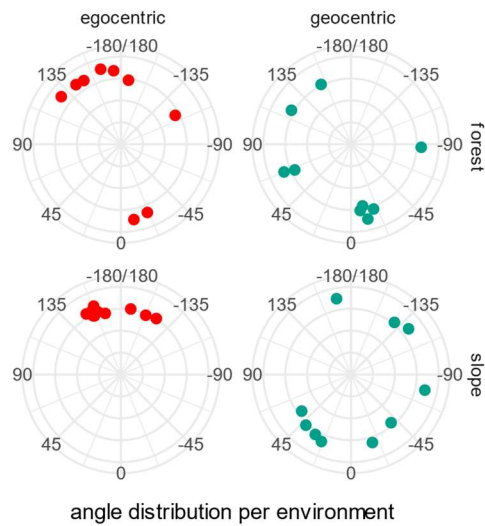


Figure 5-10: Distribution of movement angles associated with the purple sign for a single dyad (pair 17) playing the forest-slope condition.

Table 5-5: Circular sample statistics for pair 17 using the purple sign in the forest environment (block 1).

Angle relative to	$\bar{\theta}$	\bar{R}	Vm	v
Egocentric viewpoint	169.35°	0.43	0.57	74.57°
Up-down axis	-69.02°	0.40	0.6	77.79°
Across axis	20.98°	0.40	0.6	77.79°

Table 5-6: Circular sample statistics for pair 17 using the purple sign in the slope environment (block 2).

Angle relative to	$\bar{\theta}$	\bar{R}	Vm	v
Egocentric viewpoint	-6.75°	0.98	0.02	12.5°
Up-down axis	-92.15°	0.18	0.82	105.94°
Across axis	-2.15°	0.18	0.82	105.94°

Taken together, the mean resultant length is a good approximation for inferring the underlying FoR of a signal based on angle spread, although some inaccuracies may result from outliers that stray far from the mean, particularly when the data is scarce. The full dataset containing circular sample statistics for all pairs is available at https://osf.io/4azur/?view_only=3d977abf51964087a1aa61c2f28dd233.

Iconicity

The questionnaire suggested that certain signs were perceived as more iconic than others insofar as several pairs used the same sign to express the same meaning (see section 5.5.2). We can corroborate this

qualitative observation by visualising all angles produced for a given sign across all pairs and using between-dyad agreement as a proxy for iconicity. Looking at the density of movement angles, three signs stand out in relation to egocentric strategies that seemed to overlap strongly across pairs:

As discussed above, the most prevalent convergent strategy was to use the triangular signs to express *forward* and *backward*. Particularly, using the egocentric forward direction seemed to be a favourable strategy among participants. This is captured by the dense cluster of angles around $\theta_{ego} = 0^\circ$ in Figure 5-11 below.

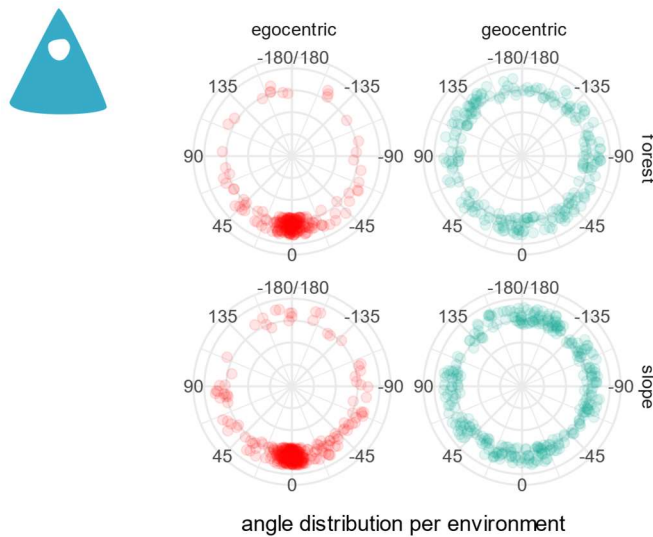


Figure 5-11: Distribution of movement angles across all participants for the turquoise sign in comparison to both FoR strategies.

For the backward direction there was slightly more noise in the associated movements, but similarly the tightest cluster of angles lies around $\theta_{ego} = \pm 180^\circ$. One pair reported using the same sign to mean forward, and several participants reported that they tried expressing backward, but it was unsuccessful. This, and the fact that the sign looks very similar to the turquoise triangle, might have caused the secondary cluster around 0° in Figure 5-12, which nevertheless lies on the same egocentric front-back axis that was often associated with these two signs.

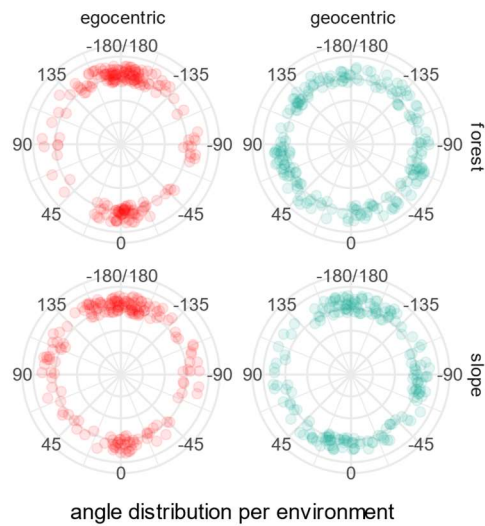


Figure 5-12: Distribution of movement angles across all participants for the turquoise sign in comparison to both FoR strategies.

Lastly, the final sign that shows a somewhat convergent movement patterns, was the grey signal. As discussed above, out of many signals used for *left* and *right*, this was the most commonly used with 16 participants reporting using it to express *right* and 8 reporting *left* in the questionnaire. This between-dyad agreement was less strong than for the triangular signals, where roughly two third of all participants agreed, but it is nevertheless somewhat reflected in the data where a clear cluster around egocentric 90° (for right) can be seen for movements in the forest environment (Figure 5-13). On the slope this cluster is slightly less dense, and the left counterpart, around -90° looks denser, compared to the forest. While this example is much noisier, it does suggest a strong tendency to exploit the left-right asymmetry of this signal.

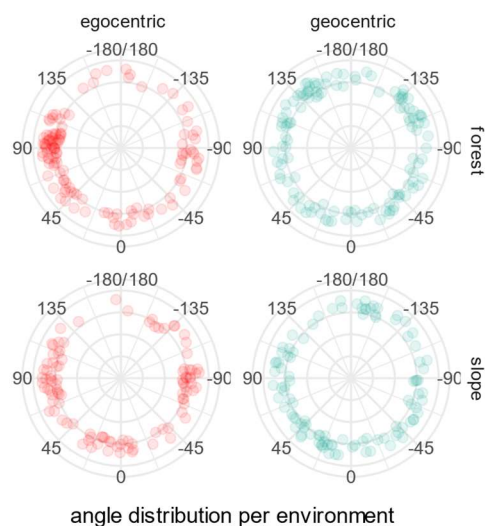


Figure 5-13: Distribution of movement angles across all participants for the turquoise sign in comparison to both FoR strategies.

Taken together, as far as iconicity is concerned, the movement data echoes observations from the questionnaire suggesting that most signals were not associated with a specific meaning. Instead, form-meaning mappings varied widely for most signals, with the exception of some convergent interpretations of the three signals discussed above. As discussed in [section 5.5.2.2](#) above, some participants did try to use the iconically designed signals to communicate geocentric directions, but this did often not work out and there was no general tendency to do so, which is why no single signal showed a strong cluster in the geocentric FoR.

Impact of environment and condition

To see which FoR strategy was more prevalent across environments and conditions, we can compare the mean resultant lengths for all signs per pair per environment for both the egocentric and geocentric FoR. For this analysis, a single mean resultant length for the geocentric FoR was obtained by calculating the mean for the up-down and across axes. In most cases the values were identical, and overall, largely correlated ($r = 0.79$) but large differences could occur, for instance, when pairs moved near the summit of the mountain slope, which shifted the subjective position of the sun and moon in the sky compared to lower areas. Furthermore, for each pair, signals were excluded if they were used less than three times in a single environment, since random sporadic movements could otherwise appear geocentric or egocentric by chance, thus introducing noise to the data. Results are shown in Figure 5-14.

The mean spread of the movement angles (the large dot) suggests that in both environments and conditions (and therefore block orders), movement strategies were on average more consistent with an underlying egocentric FoR. Additionally, it appears that egocentric strategies became even more consistent in the second block, with more angles clustering near ceiling ($\bar{R} = 1$). Likewise, there is a small visible decrease in the mean resultant length relative to the geocentric FoR in block 2.

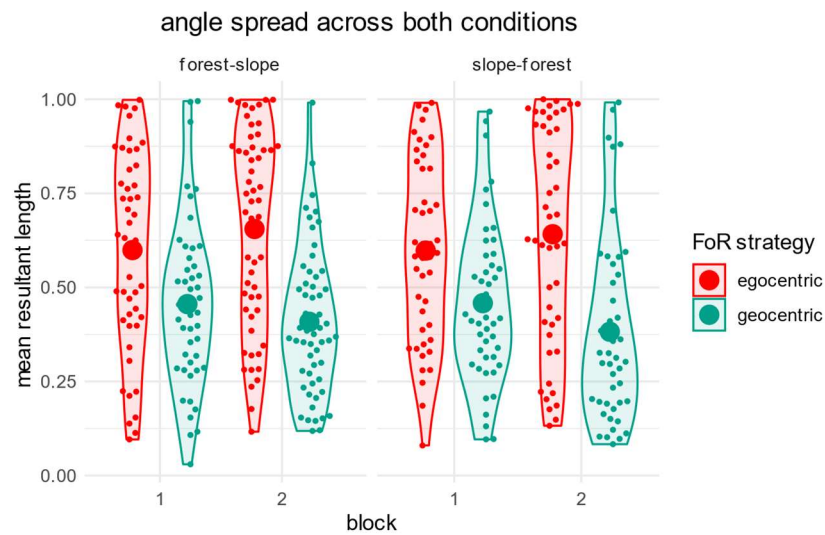


Figure 5-14: Relative spread of all angles in both experimental conditions. Points represent the mean resultant length \bar{R} of all angles associated with one sign per pair per environment in relation to the egocentric or geocentric FoR (represented by different colours).

In order to test whether FoR strategies differed between environments (slope vs forest) and block order (forest-slope condition vs slope-forest condition), a linear mixed effect model was constructed with the mean resultant length predicted by a three-way interaction between environment, block and FoR strategy. The model included random intercepts for participant pair ID and sign after a model with the maximal random effects structure did not converge. All categorical predictors (block, environment, FoR strategy) were sum coded to simplify interpretation of the intercepts. The model was fit using the `lmerTest` package (Kuznetsova, Brockhoff & Christensen 2017) package in R (R Core Team 2019).

The regression model (described in Table 5-7) yields a main effect for FoR strategy ($\beta = 0.098$, $p < 0.001$), indicating that angles associated with the egocentric FoR were in all conditions and environments much less spread out than angles associated with the geocentric FoR. Environment did not affect the choice of FoR strategy since the mean resultant length does not significantly differ between environments ($p = 0.92$). There was also no effect of condition. The order in which dyads switched between environments did not affect their FoR strategy, resulting in no significant interaction between environment, FoR strategy, and block ($p = 0.35$). There was, however, an effect of time, indicated by a significant interaction between FoR strategy and block ($\beta = -0.028$, $p < 0.05$), suggesting that the advantage of the egocentric strategy became larger over time.

Table 5-7: Fixed effects of the regression model used to analyse the mean resultant lengths of angles produced with respect to both FoR for both environments and block orders.

	Estimate	SE	<i>p</i>	
Intercept	0.523	0.018	< 0.001	***
environment (= forest)	0.001	0.012	0.926	
FoR (= egocentric)	0.098	0.012	< 0.001	***
block (= block 1)	0.003	0.012	0.834	
environment × FoR	0.011	0.012	0.361	
environment × block	0.004	0.013	0.790	
FoR × block	-0.028	0.012	0.023	*
environment × FoR × block	-0.011	0.012	0.347	

5.6 Discussion

Experiment 3 is, to my knowledge, the first experimental semiotics study that simulates the emergence of a spatial referencing system in dyadic interaction in the laboratory. While previous studies like Galantucci's (2005) language grounding game used similar referential communication games to evoke the creation of communication systems, OrbHunt is the first paradigm that allows immersing participants in a three-dimensional world with a naïve viewpoint that is central for studying spatial referencing strategies. We did not know in advance whether participants could solve the OrbHunt game, relying solely on form-meaning mappings that they had to invent and coordinate from scratch using a novel communication channel. Performance suggested that most dyads did succeed, although the game was clearly harder than in experiment 1 and 2, where language could be used to communicate.

Both a qualitative analysis based on a post-experiment questionnaire as well as a quantitative analysis based on movement data from the game suggests that there was a very strong tendency to ground symbolic meanings with respect to an egocentric FoR. Even though half of the signalling space included iconic features specifically designed to afford iconic mappings to properties of the environment, participants hardly ever conventionalized such mappings. Responses from the questionnaire revealed that some participants did try to refer to geocentric directions such as uphill or the sun, and even ad hoc landmarks, such as trees or a clearing, but they quickly abandoned it for seemingly more reliable egocentric strategies. While most signs were not associated with a particular meaning, suggesting their form was perceived as arbitrary, there was a strong tendency to use the triangular signs (that were intended to afford *uphill* and *downhill*).

In retrospect, the fact that these could easily be construed as arrows pointing forward and backward was a design oversight. However, the pilot experiment that did not use these signs had most dyads develop

signs expressing forward and backward as well. As such, the (unintentional) iconic design of these signals (that was exapted by the participants) probably did not lead to the overwhelming use of egocentric strategies; rather, the pilot and main experiment suggest that egocentric strategies are a natural consequence of the task. The navigation problem itself appears to exert a strong pressure on participants to use these directions, especially, because more fine-grained descriptions like “you went a bit too far” or distance estimates could not easily be expressed in the limited communication medium. The popularity of the triangular signs and their consistent use for these mappings then suggests that dyads simply tended to map the most important meanings on the most salient forms.

Unlike experiment 1 and 2, which used natural language, there was no difference between environments. Even on the slope, dyads favoured this egocentric strategy. This could simply reflect their cultural bias to use egocentric directions for navigation, or it could ultimately reflect a problem with the nature of the task. I speculate that if the forward/backward strategy had not been as effective as it was – which is tied to the navigational nature of the task – affordances of the environment could have played an even bigger role in grounding communication systems.

As the game moved to a more abstract experimental semiotic task in experiment 3, it became clear that the nature of the OrbHunt task is perhaps not ideal to study descriptions of spatial relations between objects. With an inventory of 8 signals, communication was much less free and participants would therefore put a lot of focus on direct navigation of the Seeker. This is effective in solving the task, but also put a lot of focus on taking the Seeker’s egocentric perspective and simply indicating the direction in which they should move with respect to their body orientation (forward, backward, left, right). This resulted in the vast majority of signs being grounded with an underlying egocentric FoR, independently of environment or block order. Ultimately, since geocentric signals were hardly used, the experiment did not really allow to see what happens to a geocentric strategy when participants switch between environments. Instead the egocentric strategy simply became more and more consistent as participants routinized and consolidated their strategy over the course of the two experimental blocks.

In a way, the results do however, speak for the flexibility of the egocentric FoR that I wanted to investigate. While we cannot generalise the findings from this task to all kinds of communities and environments, experiment 3 does suggest that in a population that has the non-linguistic egocentric FoR at their disposal, this FoR is extremely useful. An analogy would be observations made by Schlossberg (2018), who observed in his fieldwork that Marshallese speakers in Springdale, Arkansas (USA) showed a shift from geocentric to egocentric terms in their language. These urban dwellers commonly use terms corresponding to ‘left’ and ‘right’ (borrowed from English) that are rare on the Marshall Islands (see also Palmer

et al. 2017), but occur frequently in situations like driving a car around Springdale. I would suggest that an activity like driving more or less requires the presence of a non-linguistic FoR to understand relative turns to the left and right of the driver, as otherwise the driver would have to mentally represent the absolute road network (as has been suggested by Bohnemeyer et al. 2018). While surely an absolute sense of direction can be involved, it seems very costly for computing turn after turn while keeping track of absolute bearings.¹¹ Similarly, OrbHunt using signals to communicate seems to present an activity where being able to ad hoc refer to egocentric directions relative to the Seeker's looking direction is more important or useful for the task at hand. This was not the case in experiment 1 and 2, because there was no cost in simply using geocentric terms that the listener was already familiar with. However, grounding such signals from scratch turned out to be too costly and was therefore abandoned by most players in experiment 3.

Since little experimental work has previously studied spatial referencing behaviour in real-time coordination in a large-scale environments, and despite extensive testing and piloting, it was hard to predict how participants would behave in the task and what strategies they would develop. Experiment 3 did however, reveal that it is possible to simulate the emergence of an artificial spatial communication system in the laboratory and I introduced a method by which movement data extracted from the game engine can be used to directly infer the underlying spatial meanings of signals used in the game. In the following chapter, I will discuss the limitations of the OrbHunt task in more detail, how it can be improved in follow-ups as well as how future experiments using different VR tasks can be used to study the evolution and adaptation of linguistic FoR.

¹¹ However, this might be a Eurocentric perspective, since to my knowledge no driving study has compared egocentric vs allocentric driving, let alone comparing participants who rely on different FoR in their daily lives.

6 General discussion

6.1 Implications of the OrbHunt experiments

6.1.1 FoR and task-specificity

In all three OrbHunt experiments discussed in the previous two chapters, the egocentric Frame of Reference was a widespread strategy across both environments. On the one hand, this could of course reflect the linguistic sample. English speakers are highly familiar with the egocentric FoR that presents the dominant way to express spatial relations in their native language (more on this in section 6.1.2). On the other hand, it could also be a task-specific effect. OrbHunt, while designed to evoke a variety of spatial location descriptions, required a particular way of navigating and giving direction to find orbs, and thus put a lot of emphasis on navigating and the Director guiding movements of the Seeker.

From cognitive anthropology, we know that different societies engage in a wide variety of navigational practices, and often rely on very specific mental models in order to traverse their surroundings (Hutchins 1995; Thiering 2014). Certain cultural activities favour certain spatial mental models of the world. In his study on the cognitive maps used by Micronesian navigators, Gladwin (1970) describes how a navigator could not understand a simple birds-eye view map showing the locations of three islands, which Gladwin had drawn in the sand. This was because these Micronesian navigators did (at least at the time) not make use of topographic maps and would instead navigate purely relying on a sophisticated mental model that integrates natural cues (star bearings and features of wind and water) with a mental triangulation where the navigator imagines an island passing along the horizon to the side of canoe, which indicates distance to the destination in units of time (see Hutchins 1983). For this, they assume a frame of reference, where the origin of the coordinate system lies in the boat, which is imagined as static, while bodies of land would move towards and away from the navigator (Hutchins & Hinton 1984). This system serves as a perfect adaptation to the local environment and requires a certain geographical distribution of islands and visible stars in order to be effective. While Western anthropologists were puzzled by this mental skill for over a century (Schück 1882; Hutchins 1983), such an adaptation allowed Micronesian navigators – presumably for millennia – to traverse vast distances without instruments and techniques such as the log for dead-reckoning that were required for similar undertakings by Western navigators.

Such observations are relevant, because field-linguists have similarly reported that specific Frames of Reference appear to vary as a function of engagement with specific cultural activities, e.g., engagement with the sea through activities like fishing Palmer et al. (2017) or driving (Schlossberg 2018),

which can lead to dramatic differences in whether speakers of ‘atoll’-languages will employ a geocentric or egocentric FoR. Similarly, work in rural Mexico (Adamou & Shen 2016; Calderón et al. 2019) and the Peruvian Andes (Shapero 2017) have shown that small cohesive cultural groups with ways of lives that are closely tied to their surrounding natural and constructed environments and involve frequent interaction with these through cultural activity, are more likely to use spatial language that is environmentally grounding (Palmer et al. 2017 call this the Sociotopographic Correspondence Model).

However, even looking at educated, globalised, post-industrial, urban speakers, we can find similar task-dependent effects: Tenbrink (2018), for instance describes how specific activities like horse-riding and sailing can involve surprisingly specialised FoR that can differ dramatically from the prototypical scholars would usually associate with members of these communities. In sailing, true ‘forward’ movement is hardly possible due to forces acting on the boat requiring specific goal-directed spatial reasoning and long-term planning. Sailing terminology, which is highly spatial, reflects directional concepts that are normally not encountered in everyday life (see Tenbrink & Dylla 2017 for details). Similarly, in circular horse-riding *inside* and *outside* and the ad hoc integration of landmarks become more prominent than egocentric *left/right*, which would, e.g., be used in driving by members of the same community (Tenbrink 2018).

In light of this evidence, just like it is problematic to treat languages as single datapoints (see section [3.2](#)), no single spatial referencing task can ever uncover ‘how speakers of this language do it’. Therefore, we have to assume similar task-specific effects for the OrbHunt game. Schlossberg (2018) reports that Marshallese speakers in the US use an egocentric perspective when driving. This makes a lot of sense, given that driving focuses heavily on turning left and right as well as forwards and backwards. The OrbHunt task is quite similar in this regard, except for the fact that it also involves perspective-taking and locating external objects. OrbHunt can therefore only provide a full picture of spatial referencing strategies in this type of navigational task. One advantage of VR is that even assuming that such task-specific effects will occur, it allows for controlling essentially everything that happens in the task. Experiment 1 and 2 show that even for the same referential task, manipulating only the environment, can have a significant effect on FoR descriptions in a sample of English speakers. This adds further evidence to the flexibility and environmental sensitivity of human FoR strategies that is also reflected in the plurality and diversity of spatial cultural practices discussed above. Using the same environments in additional experiments with different tasks (e.g., describing relations between many small objects) could then be used to generalise the findings concerning the impact of these environments. OrbHunt currently only allows making very general claims about the impact of the topographic environment on spatial language, but

additional experiments combining a larger selection of carefully controlled environments and different referential task could be used to test more explicit claims on the impact of very specific environments on spatial language (e.g., slopes, atolls, rivers etc.; see also the following section on potential follow-ups).

Additionally, the results from experiment 3 raise the question of why this environmental effect disappeared after the medium of communication was changed. There are various possibilities. Firstly, the need to ground a symbolic communication system made the task much harder. In order to collect orbs in the same short amount of time, participants had to conventionalise efficient and effective strategies to solve this task, while feedback existed only in the form of trial and error. Since every signal-to-meaning mapping had to be negotiated in this way, there was a pressure to create simple systems that could do with a few reliable signals. Forward systems turned out to be extremely useful for this, and the tendency to create them was further boosted by exapting the iconicity of triangular shapes and using them as indices. Perhaps this made experiment 3 feel even more like a navigation task rather than a game about ‘communicating locations’. Secondly, it is possible that the predominant strategy in participants’ native language (egocentric across both environments in experiment 1 and 2), was amplified further by the pressure for creating a novel communication system. While in experiment 1 and 2, participants could easily rely on their native language’s repertoire for describing spatial locations, the very restricted affordances of experiment 3 required them to coordinate the most salient and effective solution. In experiment 3, it may have been easier to provide feedback when the Seeker successfully re-orientes with respect to an egocentric intrinsic FoR. Directions with respect to one’s own body may be easier to track than geocentric ones if this cognitive style is usually preferred in everyday life.

Lastly, while VR setups like the OrbHunt game can be said to provide increased ecological validity by immersing participants in realistic settings, the task itself cannot be said to be completely naturalistic. Rather, OrbHunt can only be said to investigate spatial referencing pertaining to one specific activity that should be expected to motivate specific solutions (similar to how activities like driving or sailing would). In the present task, the focus on navigation clearly privileged an egocentric solution, particularly, in the experimental semiotics variant of experiment 3, where a simple communication system consisting of *forward* and *turn left/right* signals was extremely effective in solving the task. Similarly, in experiment 2 and 3 there were direct consequences to participants changing environments. For instance, when changing from slope to forest, the affordance for grounding an *uphill/downhill* system disappeared and if such a system was pre-established, it had to be abandoned. In this way, both the design of the referential communication game and the affordances of the environment determine the range of strategies that can be observed in the lab. Additionally, some aspects of the design, such as the limited visibility of the Director and

orbs from the Seeker's perspective were not very realistic. It is therefore important to combine this experimental approach with qualitative and descriptive fieldwork that surveys the full range of activities involving spatial referencing and the possible variables influencing them (see also [Chapter 3](#)). A series of experiments can then systematically investigate these variables and activities in more detail. Some potential future experiments that should complement the findings from the present experiments are discussed in the following section.

6.1.2 Potential follow-ups and future experiments

Populations and environments

Task-specific effects naturally interact with language-specific effects and in the future, the OrbHunt game should be conducted with speakers from other linguistic and cultural communities. It is, e.g., unclear how speakers from predominantly geocentric communities would solve the task: In experiment 3, would they overwhelmingly rely on the geocentric FoR? Or would they align with the egocentric result from the English speakers I tested? Analogously, the task could be run with members from communities that contrast on other sociotopographic levels such as urban vs rural (e.g., living in the mountainous areas featuring slopes), different levels of engagement with navigational activities (driving, sailing etc) or which FoR they prefer linguistically vs non-linguistically.

As far as environments are concerned, the geometric analysis developed for experiment 3 showed that movements in absolute directions and relying on a geocentric FoR can readily be distinguished from movements based on an egocentric FoR. For future experiments, it would therefore make sense to afford absolute directions even more clearly. For example, by placing salient and fixed landmarks (such as coloured towers) in all four cardinal directions, so that these can be reliably used to track these directions. Besides 'natural' environments, future tasks should also look at large-scale urban environments (e.g., a coastal town with nuclear structures vs a grid-like city).

Tasks and scales

The OrbHunt game is a proof-of-concept that shows that VR experiments can be used to study the production of spatial language in realistic large-scale environments to test very specific hypotheses. This means, VR can complement table-top tasks, photographs of objects and the general tendency to conduct experiments in small enclosures, which might affect responses in spatial linguistic tasks. This is particularly, relevant when studying the relation of spatial language to the environment, since humans move outside, in large-scale environments, and it is also relevant in thinking about the prehistory of spatial

referencing strategies, as we must have used spatial language before we widely settled down and began introducing more and more small manipulable material artifacts.

Since OrbHunt ended up eliciting many descriptions focusing on ‘directing’ the Seeker, future referential communication games should be refined to focus more on communicating relations between objects.¹² OrbHunt showed how in VR, arrays of objects can be controlled even with respect to which objects are visible to whom. One could think of large-scale VR version of the Man and Tree Space (e.g., Haun et al. 2011), where rather than on a table, participants can move between larger objects in various environments with different affordances for grounding geocentric language (e.g., a river, a slope etc, landmarks etc). With respect to scales, one could even look at transfer between scales, e.g., if participants are first solving a large-scale task, and are subsequently presented with miniature versions of the same objects they encountered. If such a setting is done in an experimental semiotics paradigm, where spatial vocabulary has to be grounded first, one could study how spatial descriptions evolve if they have to be applied to a different scale (e.g., to test whether egocentric language is more adapted for talking about small manipulable objects, as speculated by Bohnemeyer 2017).

Similarly, VR allows creating previously unthinkable tasks, involving zero-gravity, mismatches in scale (e.g. ,shrinking one participant with respect to the other), circular space, walking on walls etc. Such unnatural scenarios could be used to fully test the human ability for spatial referencing at its limits, study novel coordinate systems or techniques that are normally not encountered in the real world and see how existing referencing strategies can be extended and applied to completely novel situations. This is relevant, since spatial language will keep evolving with our niche-constructing species that might eventually inhabit rotating space stations (Messerschmid & Bertrand 2013) and is already today facing new challenges, e.g., in describing the intricately complex spatial dimensions of folded proteins that are highly relevant to medicine medical applications (e.g., Onuchic & Wolynes 2004). However, most importantly such studies would add to the emerging picture that almost every year, new findings seem to suggest that human spatial referencing is more flexible and diverse than previously assumed.

Modelling evolution

An evolutionary, causal approach, such as the one laid out in [Paper 3](#), seems particularly relevant here, since it does not try to reduce the complexity of the issue. Instead, observational data is used to formulate

¹² Although it is worth mentioning that the style in which directions are given could also be varied. When giving someone directions, one would usually not follow them, but rather provide several descriptions (*turn left, then right* . .). A more bee-dance style game, similar to the Virtual Atoll Task (Lum & Schlossberg 2014) could be an alternative.

careful and informed hypotheses that can then step-by-step be tested experimentally, to gradually complete a mechanistic mosaic that accounts for all variables involved and can model the actual diversity we observe.

While in this thesis, I have focused on the emergence of conventions in face-to-face interactions, cultural evolution does of course also involve transmission, which has likewise been studied by linguists looking at spatial language (e.g., Bohnemeyer et al. 2015; Meakins et al. 2016; Calderón et al. 2019). Future experiments should therefore make use of the Iterated Learning paradigm and simulate how different FoR strategies evolve as they are passed on from one population to the next. It would be very valuable to show experimentally, not only that different strategies emerge in different environmental settings, but also that these strategies can then ‘grammaticalize’ and remain invariant even if new environments are encountered. In contrast to experiment 3, one could also explicitly teach participants artificial languages that already encode specific FoRs and see how these are retained or adapted in different environments, or what happens if participants encounter a partner trained on a different system (which strategy will prevail after which condition?).

Future, experimental semiotics paradigms like experiment 3 could benefit from improved stimuli, since at the design stage, it was very difficult to judge how participants would use the visual signals. Instead, one could use ‘iconicity ratings’, i.e. having a large number of native speakers rate the iconicity of many visual signs for expressing specific spatial relations, similar to ratings used to study sound symbolism in lexical items (e.g., Winter et al. 2017). This could reveal which types of shapes would be particularly well-suited to train participants on a specific FoR, which could then subsequently be confronted with different conditions to study how these affect the FoR strategy (different environments, scales, referential tasks, contact situations etc).

Finally, these VR tasks are not limited to the study of language, but can also be used to study body orientation and co-speech gestures that have been of interest in spatial referencing research (e.g., Haviland 1993; Haun & Rapold 2009; Meakins, Jones & Algy 2016; Calderón, De Pascale & Adamou 2019), since VR allows tracking head movements, looking directions, and body orientations in real-time if the right controllers and trackers are used (see, e.g., Peeters 2019).

7 Concluding remarks

7.1 Recapitulation

At the beginning of this thesis, I asked: why are there different languages? I referred to increasing evidence which points to adaptation as at least one possible mechanism shaping cross-linguistic diversity. In [Chapter 1](#), I set the scene by briefly introducing the field of language evolution research (Nölle, Hartmann & Tinitis 2020), and reviewing findings that suggest that language evolves not in isolation, but rather situated and in interaction with the biocultural and physical niches it inhabits. The two main research projects I presented set out to investigate linguistic adaptation from a cultural evolutionary perspective in the laboratory, each of them focusing on different aspects.

The silent gesture experiments presented in [Chapter 2](#) investigated whether systematicity – a property fundamental to all human communication systems (be it signed or spoken language, morse codes etc) – would adapt to external pressures. We found that the referential environment in which novel gestural communication systems emerged could indeed affect which categories would become systematized as well as the extent to which participants conventionalised systematic markers and how stable these remained over time (Nölle et al. 2018b). These results suggest that environmental factors can shape the evolution of linguistic structure even at a very fundamental level, and the finding that statistical relationships of traits in the environment are reflected in the categories that emerged, might help explain why different cultures end up with different category systems to fulfil their communicative needs.

From [Chapter 3](#), I then turned to spatial language as a more concrete phenomenon at the intersection of language, cognition and the environment. I reviewed research showing that variation in spatial Frames of Reference, is one of the most complex cases of linguistic features being deeply intertwined with cognition, and embedded in highly situated culture-specific activities (e.g., Thiering 2014). An evolutionary, causal approach was presented to untangle this problem by combining traditional descriptive methods with models and experiments that can test concrete causal hypotheses and uncover underlying mechanisms (Nölle & Spranger accepted).

[Chapter 4](#) then presented a series of VR experiments that tested the specific hypothesis that spatial Frames of Reference are sensitive to topographic affordances. Experiment 1 found that English-speaking dyads playing the same spatial referential communication game ('OrbHunt') in different environments (a forest and a slope) would attune their spatial location descriptions with respect to these environments, thus producing more geocentric descriptions on the slope. Experiment 2 addressed whether dyads switching environments would flexibly switch or retain their FoR strategy. The experiment replicated findings from

experiment 1, but contrary to my predictions there was no difference between two different orders in which they encountered the forest and slope environments. This confirms that speakers of languages that support multiple FoR, can readily and flexibly switch between these FoR in response to the environment at hand.

[Chapter 5](#) then added a third experiment using an experimental semiotics approach, where participants had to solve the same OrbHunt task, but could only use an artificial communication medium to communicate locations and directions. This was done, because I expected differences in flexibility between egocentric and geocentric strategies across environments, which show that there is a cost to switching strategies (in this case, the cost to re-negotiate communicative conventions through trial and error). However, in contrast to experiments 1 and 2, dyads in experiment 3 established predominantly egocentric strategies across the board, and there were no differences between environments.

Nevertheless, experiment 3 illustrated how VR allows studying spatial referencing on the large scale in a controlled environment and was to my knowledge the first lab experiment that had participants successfully ground complex spatial referencing systems from scratch. Furthermore experiment 3 presented a novel methodology for using movement data to infer underlying Frames of Reference in novel signals.

This results in a lot of possibilities for using VR and behavioural measures of this kind to study spatial language in controlled large-scale settings in the lab. In [Chapter 6](#), I discuss several possibilities for future experiments and direct follow-ups that address some of the limitations of the OrbHunt paradigm presented in this thesis.

7.2 Final thoughts

When I began this PhD project, commercial VR technology was completely new, and it took me a year to simply set up the lab and begin designing the setups for the experiments reported here. Since then, things have rapidly changed, linguists have begun to see its potential (see, e.g., Peeters 2019), and more user-friendly APIs and development kits have made VR more accessible than ever. With the available technology, I was not only able to use VR to immerse participants in large-scale environments, but also hook up two computers to the same world, where the experiments took place. For studying communication in its full multimodal richness under controlled conditions, this possibility is a game-changer and has a huge potential for language evolution research.

I therefore see the contribution of this thesis not only in presenting novel evidence for languages adapting to their environment, but also in the methodological innovations. I believe that an evolutionary perspective that tries to model causal relationships and their underlying mechanisms is necessary when

dealing with messy and complex cultural evolutionary phenomena, such as language and its relation to other aspects of human life.

Appendices

Appendix A: Data and code experiments 1-3

All coded dialogues as well as CSV files containing the movement data and looking directions for the experiments 1-3 (see chapters 4 and 5) as well as the responses from the questionnaire (experiment 3) are stored in an OSF repository and available under the following link:

https://osf.io/4azur/?view_only=3d977abf51964087a1aa61c2f28dd233

Appendix B: OrbHunt: Setup and technical specifications

Technical specifications: OrbHunt was built with the Unity 3D engine (version 2017.3, see <https://unity.com/>). The experiments were run on two identical HP ENVY 750 computers with the following specifications:

Graphics: Radeon™ RX 480 with 1x HDMI and 3x DisplayPort outputs

Processor: Intel® Core™ i5-6400

Memory: 8 GB DDR4 SDRAM

Storage: 128 GB solid-state drive and 1 TB hard drive

USB ports: 4x USB 2.0 and 4x USB 3.0

Operating system: Windows 10 Home

Mouse and keyboard: USB wired keyboard and optical mouse

Optical drive: Ultra Slim-Tray SuperMulti DVD Burner

Wireless: Wireless LAN 802.11a/b/g/n/ac and Bluetooth® 4.0 M.2 combo

Each computer was equipped with an HTC Vive head-mounted device (1st generation) and two wireless handheld controllers. The base stations (for tracking the headsets and controllers) were mounted on poles in diagonally opposite corners of the room and

Setup: The entire OrbHunt Unity project can be found at the following GitHub repository: <https://github.com/Jonas-N/OrbHuntSetup>.

The directory includes all project files necessary to open the game in Unity 2017.3. This version of OrbHunt is the version used for experiment 3, which allows creating builds with the functionality needed to run all three experiments, e.g., by disabling the symbol communication channel and the environment switch after round 5 of the game.

Inside the main project folder, the `Assets` folder contains all assets used to create the OrbHunt game. There are folder for the shaders, materials terrain information and textures that appear in the game. The folder `Imports` includes all external APIs used for the setup, such as Valve's SteamVR API, which was used for the VR components, and the Photon API, used to run the experiment over network to connect both computers. The `Prefabs` folder contains a `Resources` folder that includes all prefab game objects that become instantiated on the network (such as the Seeker, Director, orbs, orb effects etc). The `Scenes` folder includes 1) The Launcher scene that allows the experimenter to enter participant data, choose conditions and start the game, 2) the training environment that is initially loaded, and 3) the forest and slope

scenes as well as two separate folders with the lighting data for both scenes. Finally, the most important folder is the `Scripts` folder, which contains 31 original C# scripts that are attached to the various game objects in the scenes and prefabs that become instantiated in the network in order to run the game. All of the scripts are extensively commented and allow changing parameters such as the round time, number of rounds, orb spawn distance etc.

Appendix C: Instructions for experiments 1-3

Note that these instruction protocols were delivered verbally to the participants by the experimenter, since participants were wearing their VR headsets while being instructed (to explain them the game interface and controls). While participants could ask clarification questions, the experimenter consciously avoided spatial language, so as not to prime participants with suggestions for how to solve the task.

Experiment 1

Welcome to this VR experiment!

Before we start, I just want to make sure that this will be a comfortable experience. If you feel uncomfortable during the experiment for any reason, please inform me. If you want to stop the experiment, remove your headset.

In this experiment, you will play a coordination game with a partner. In the game you have to work together on collecting orbs to score points. For this you will have to communicate with your partner. We will record your spoken conversation for our data analysis.

[Training environment]

These are your controllers. By clicking the **touchpad**, you can teleport to move around from point to point. Try it out - you will teleport to where the curve points.

In the actual game, you will also need to grab orbs, for this you can either use the **trigger** at the back of your controllers or the **grip buttons** at the side of each controller's grip (easier). You will also start new rounds with the trigger.

Before we start let me briefly explain the game. The game consists of **10 rounds of 2 minutes** each. Every round, one of you will be the **Seeker** and the other one will be the **director**.

The task is to find yellow **orbs** that give you **points**. This is quite hard for the seeker, as they can only see the orb when they get very close to it. However, the director can see orbs from afar (there is even a light pillar indicating its location) and can thus direct the seeker. The director can also see the Seeker, but not vice versa. Therefore, you will need to communicate to find the orbs.

Each round the roles switch. When you are ready, I will start the game. Good luck!

Experiment 2

Welcome to this VR experiment!

Before we start, I just want to make sure that this will be a comfortable experience. If you feel uncomfortable during the experiment for any reason, please inform me. If you want to stop the experiment, remove your headset.

In this experiment, you will play a coordination game with a partner. In the game you have to work together on collecting orbs to score points. For this you will have to communicate with your partner. We will record your spoken conversation for our data analysis.

[Training environment]

These are your controllers. By clicking the **touchpad**, you can teleport to move around from point to point. Try it out - you will teleport to where the curve points.

In the actual game, you will also need to grab orbs, for this you can either use the **trigger** at the back of your controllers or the **grip buttons** at the side of each controller's grip (easier). You will also start new rounds with the trigger.

Before we start let me briefly explain the game. You will start in a forest/on a mountain. The game consists of **10 rounds of 2 minutes** each. After 5 rounds you will **switch** to another environment. Every round, one of you will be the **Seeker** and the other one will be the **director**.

The task is to find yellow **orbs** that give you **points**. This is quite hard for the seeker, as they can only see the orb when they get very close to it. However, the director can see orbs from afar (there is even a light pillar indicating its location) and can thus direct the seeker. The director can also see the Seeker, but not vice versa. Therefore, you will need to communicate to find the orbs.

Each round the roles switch. When you are ready, I will start the game. Good luck!

Experiment 3

Welcome to this VR experiment!

Before we start, I just want to make sure that this will be a comfortable experience. If you feel uncomfortable during the experiment for any reason, please inform me. If you want to stop the experiment, remove your headset.

In this experiment, you will play a coordination game with a partner. In the game you have to work together on collecting orbs to score points. For this you will have to communicate with your partner using symbols. These symbols have no inherent meaning. You will have to negotiate the meaning of the symbols yourself through trial and error. It seems daunting at first, but you will find that usually after a few rounds you will be able to communicate successfully.

[Training environment]

These are your controllers. By clicking the **touchpad**, you can teleport to move around from point to point. Try it out - you will teleport to where the curve points.

In the actual game, you will also need to communicate using symbols for which you use the **grip buttons** to select symbols, and the **trigger** to send them. You will also start new rounds using the **trigger**.

Before we start let me briefly explain the game. You will start in a forest/on a mountain. The game consists of **10 rounds of 2 minutes** each. After 5 rounds you will **switch** to another environment. Every round, one of you will be the **Seeker** and the other one will be the **director**.

The task is to find yellow **orbs** that give you **points**. This is quite hard for the seeker, as they can only see the orb when they get very close to it. However, the director can see orbs from afar (there is even a light pillar indicating its location) and can thus direct the seeker. The director can also see the Seeker, but not vice versa. Therefore, you will need to communicate to find the orbs. Please only use symbols to communicate and do not speak (except to the experimenter if you have questions).

Each round the roles switch. When you are ready, I will start the game. Good luck!

Appendix D: Post-test questionnaire for experiment 3

The questionnaire instructed participants as follows:

Please describe how you used the symbols to communicate.

For each symbol that you used, please provide:

- 1) a short description such as "used to communicate that my partner should turn around" or "first used when the orb was left of my partner, but in the second environment we changed the meaning to..."
- 2) a simple label/phrase that expresses the symbol's meaning such as "left" or "downhill"

If you haven't used a symbol, just leave the fields empty.

For each symbol participants then had to answer two questions:

"How did you use this symbol? Please also mention meaning changes over time/between environments."

"How would you label it? (simple phrase or word)"

Finally, they could leave a comment on what worked well/didn't work so well in the task and what strategy they used, as well as information on the languages they speak. The full questionnaire can be accessed at: <https://docs.google.com/forms/d/1WTod6whVmRVoOoSo1IWS86etdNdLqtJYNPjqRWD5QQ/edit>

References

- Adamou, Evangelia & Stefano De Pascale. submitted. A meta-analytic study of the ‘Animals in a Row’ task: Assessing the impact of spatial language on small-scale memorization. *Linguistics Vanguard*.
- 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.
<https://doi.org/10.1163/15685373-12342193>.
- Agostinelli, Claudio & Ulric Lund. 2017. R package “circular”: Circular Statistics (version 0.4–93). *CA: Department of Environmental Sciences, Informatics and Statistics, Ca’ Foscari University, Venice, Italy. UL: Department of Statistics, California Polytechnic State University, San Luis Obispo, California, USA*. <https://cran.r-project.org/web/packages/circular/>.
- Allen, Gary L. 1999. Spatial abilities, cognitive maps, and wayfinding. *Wayfinding behavior: Cognitive mapping and other spatial processes* 4680.
- Ameka, Felix K. & J. Essegbey. 2006. Elements of the grammar of space in Ewe. In *Grammars of space: Explorations in cognitive diversity*, 359–399. Cambridge University Press.
- Ammon, Ulrich, Norbert Dittmar, Klaus J. Mattheier & Peter Trudgill (eds.). 2004. *Sociolinguistics: An International Handbook of the Science of Language and Society*. 2nd edn. Vol. 1. Berlin, New York: Walter de Gruyter.
- Aoki, Kenichi & Marcus W. Feldman. 1991. Recessive hereditary deafness, assortative mating, and persistence of a sign language. *Theoretical Population Biology* 39(3). 358–372.
[https://doi.org/10.1016/0040-5809\(91\)90029-F](https://doi.org/10.1016/0040-5809(91)90029-F).
- Arbib, Michael A. 2012. *How the brain got language: The mirror systems hypothesis*. Oxford: Oxford University Press.
- Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68. 255–278.
- Barsalou, Lawrence W. 1999. Perceptual symbol systems. *Behavioral and Brain Sciences* 22(04).
<https://doi.org/10.1017/S0140525X99002149>.
- Barsalou, Lawrence W. 2008. Grounded Cognition. *Annual Review of Psychology* 59(1). 617–645.
<https://doi.org/10.1146/annurev.psych.59.103006.093639>.
- Bartlett, Frederic C. 1932. *Remembering. A study in experimental and social psychology*. Cambridge: Cambridge University Press.
- Bates, Douglas, Reinhold Kliegl, Shravan Vasishth & Harald Baayen. 2015. Parsimonious Mixed Models. *arXiv:1506.04967 [stat]*. <http://arxiv.org/abs/1506.04967> (29 June, 2018).
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1). 1–48.
<https://doi.org/10.18637/jss.v067.i01>.
- 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.
<https://doi.org/10.1111/j.1467-9922.2009.00533.x>.
- Bentz, Christian, Dan Dediu, Annemarie Verkerk & Gerhard Jäger. 2018. The evolution of language families is shaped by the environment beyond neutral drift. *Nature Human Behaviour* 2(11). 816–821. <https://doi.org/10.1038/s41562-018-0457-6>.
- Bentz, Christian & Bodo Winter. 2013. Languages with More Second Language Learners Tend to Lose Nominal Case. *Language Dynamics and Change* 3(1). 1–27.
<https://doi.org/10.1163/22105832-13030105>.

- Bergen, Benjamin K. 2004. The Psychological Reality of Phonaesthemes. *Language*. Linguistic Society of America 80(2). 290–311. <https://doi.org/10.1353/lan.2004.0056>.
- Berwick, Robert C. & Noam Chomsky. 2016. *Why Only Us: Language and Evolution*. MIT Press.
- Berwick, Robert C. & Noam Chomsky. 2017. Why only us: Recent questions and answers. *Journal of Neurolinguistics* 1–12. <https://doi.org/10.1016/j.jneuroling.2016.12.002>.
- Bickel, Balthasar. 1999. Cultural formalism and spatial language in Belhara. *Himalayan space: Cultural horizons and practices*. Voelkerkundemuseum der Universitaet Zürich Zürich 75–104.
- Bickel, Balthasar. 2015. Distributional typology: Statistical inquiries into the dynamics of linguistic diversity. In Bernd Heine & Heiko Narrog (eds.), *Oxford Handbook of Linguistic Analysis, 2nd edition*, 901–923. Oxford: Oxford University Press.
- Bickerton, Derek & others. 2009. *Adam's tongue: how humans made language, how language made humans*. Macmillan.
- Blackstock, J. & J. Brown. 2009. *Bitstrips*. Toronto: Canada.
- Blasi, Damián E., S. Moran, S. R. Moisik, P. Widmer, D. Dediu & B. Bickel. 2019. Human sound systems are shaped by post-Neolithic changes in bite configuration. *Science* 363(6432). eaav3218. <https://doi.org/10.1126/science.aav3218>.
- 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.
- Blasi, Damián E., Søren Wichmann, Harald Hammarström, Peter F. Stadler & Morten H. Christiansen. 2016. Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences* 113(39). 10818–10823. <https://doi.org/10.1073/pnas.1605782113>.
- Blevins, Juliette. 2004. *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge: Cambridge University Press.
- Bohil, Corey J., Bradly Alicea & Frank A. Biocca. 2011. Virtual reality in neuroscience research and therapy. *Nature Reviews Neuroscience* 12(12). 752–762. <https://doi.org/10.1038/nrn3122>.
- Bohnenmeyer, Jürgen. 2011. Spatial frames of reference in Yucatec: Referential promiscuity and task-specificity. *Language Sciences* 33(6). 892–914. <https://doi.org/10.1016/j.langsci.2011.06.009>.
- Bohnenmeyer, Jürgen. 2016. The Interplay of Language, Culture, and Environment in Spatial Reference. In. Santa Barbara.
- Bohnenmeyer, Jürgen. 2017. The conquest ^[1]_{SEP} of manipulable space. In, 55. Australian National University.
- Bohnenmeyer, Jürgen, Katharine Donelson, Yen-Ting Lin & Randi Moore. 2018. The biological and cultural evolution of spatial cognition: Evidence from semantic typology. In. Rome, Italy.
- Bohnenmeyer, 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.
- Bohnenmeyer, Jürgen & Carolyn O'Meara. 2012. Vectors and frames of reference: Evidence from Seri and Yucatec. In Luna Filipović & Kasia M. Jaszczolt (eds.), *Space and Time in Languages and Cultures*, 217–249. John Benjamins.
- Botha, Rudolf. 2012. Inferring modern language from ancient objects. In Maggie Tallerman & Kathleen R. Gibson (eds.), *The Oxford Handbook of Language Evolution*, 303–312. Oxford: Oxford University Press.
- Bouckaert, Remco R., Claire Bowern & Quentin D. Atkinson. 2018. The origin and expansion of Pama–Nyungan languages across Australia. *Nature Ecology & Evolution* 2(4). 741–749. <https://doi.org/10.1038/s41559-018-0489-3>.

- Bowerman, Melissa & Soonja Choi. 2001. Shaping meanings for language: universal and language-specific in the acquisition of spatial semantic categories. In Melissa Bowerman & Stephen Levinson (eds.), *Language Acquisition and Conceptual Development* (Language Culture and Cognition), 475–511. Cambridge: Cambridge University Press.
<https://doi.org/10.1017/CBO9780511620669.018>.
- Bozgeyikli, Evren, Andrew Raji, Srinivas Katkooori & Rajiv Dubey. 2016. Point & Teleport Locomotion Technique for Virtual Reality. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (CHI PLAY '16), 205–216. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/2967934.2968105>.
- Brewer, Mark J., Adam Butler & Susan L. Cooksley. 2016. The relative performance of AIC, AIC_c and BIC in the presence of unobserved heterogeneity. (Ed.) Robert Freckleton. *Methods in Ecology and Evolution* 7(6). 679–692. <https://doi.org/10.1111/2041-210X.12541>.
- Brezina, Vaclav. 2018. *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge University Press. <https://doi.org/10.1017/9781316410899>.
- Brighton, Henry. 2002. Compositional Syntax From Cultural Transmission. *Artificial Life*. MIT Press 8(1). 25–54. <https://doi.org/10.1162/106454602753694756>.
- 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.
<https://doi.org/10.1017/S0952523804213098>.
- Brown, Penelope. 2006. A sketch of the grammar of space in Tzeltal. *Grammars of space: Explorations in cognitive diversity*. Cambridge University Press 230–272.
- Brown, Penelope & Stephen C. Levinson. 1993. “Uphill” and “Downhill” in Tzeltal. *Journal of Linguistic Anthropology* 3(1). 46–74. <https://doi.org/10.1525/jlin.1993.3.1.46>.
- Burnham, Kenneth P & David R Anderson. 2004. Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research*. Sage Publications Sage CA: Thousand Oaks, CA 33(2). 261–304.
- Bybee, Joan L. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
- 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.
<https://doi.org/10.1016/j.langsci.2019.04.001>.
- Caldwell, Christine A. & A. E Millen. 2008. Studying cumulative cultural evolution in the laboratory. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363(1509). 3529–3539.
<https://doi.org/10.1098/rstb.2008.0133>.
- Caldwell, Christine A. & Kenny Smith. 2012. Cultural Evolution and Perpetuation of Arbitrary Communicative Conventions in Experimental Microsocieties. *PLOS ONE* 7(8). e43807.
<https://doi.org/10.1371/journal.pone.0043807>.
- Caramazza, Alfonso & Bradford Z. Mahon. 2006. The organisation of conceptual knowledge in the brain: The future’s past and some future directions. *Cognitive Neuropsychology* 23(1). 13–38.
<https://doi.org/10.1080/02643290542000021>.
- Caramazza, Alfonso & Jennifer R. Shelton. 1998. Domain-Specific Knowledge Systems in the Brain: The Animate-Inanimate Distinction. *Journal of Cognitive Neuroscience* 10(1). 1–34.
<https://doi.org/10.1162/089892998563752>.
- Carlson-Radvansky, Laura A. & David E. Irwin. 1993. Frames of reference in vision and language: Where is above? *Cognition* 46(3). 223–244. [https://doi.org/10.1016/0010-0277\(93\)90011-J](https://doi.org/10.1016/0010-0277(93)90011-J).
- Carr, Jon W., Kenny Smith, Hannah Cornish & Simon Kirby. 2017. The Cultural Evolution of Structured Languages in an Open-Ended, Continuous World. *Cognitive Science* 41(4). 892–923.
<https://doi.org/10.1111/cogs.12371>.

- Carr, Jon W., Kenny Smith, Jennifer Culbertson & Simon Kirby. 2020. Simplicity and informativeness in semantic category systems. *Cognition* 202. 104289. <https://doi.org/10.1016/j.cognition.2020.104289>.
- Carroll, Mary. 1993. Deictic and Intrinsic Orientation in Spatial Descriptions: A Comparison Between English and German. In Jeanette Altarriba (ed.), *Advances in Psychology* (Cognition and Culture), vol. 103, 23–44. North-Holland. [https://doi.org/10.1016/S0166-4115\(08\)61655-6](https://doi.org/10.1016/S0166-4115(08)61655-6).
- Carstensen, Alexandra, Jing Xu, Cameron Smith & Terry Regier. 2015. Language evolution in the lab tends toward informative communication. In *CogSci*.
- Cassidy, Kimberly Wright & Michael H. Kelly. 1991. Phonological information for grammatical category assignments. *Journal of Memory and Language* 30(3). 348–369. [https://doi.org/10.1016/0749-596X\(91\)90041-H](https://doi.org/10.1016/0749-596X(91)90041-H).
- Castillo, Lucía, Kenny Smith & Holly P. Branigan. 2019. Interaction Promotes the Adaptation of Referential Conventions to the Communicative Context. *Cognitive Science* 43(8). e12780. <https://doi.org/10.1111/cogs.12780>.
- Chekaf, Mustapha, Nelson Cowan & Fabien Mathy. 2016. Chunk formation in immediate memory and how it relates to data compression. *Cognition* 155. 96–107. <https://doi.org/10.1016/j.cognition.2016.05.024>.
- Chen, M. Keith. 2013. The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets. *American Economic Review* 103(2). 690–731.
- Choi, Edward, Angeliki Lazaridou & Nando de Freitas. 2018. Compositional Obverter Communication Learning From Raw Visual Input. *arXiv:1804.02341 [cs]*. <http://arxiv.org/abs/1804.02341> (28 October, 2020).
- Choi, Soonja & Melissa Bowerman. 1991. Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition* 41(1–3). 83–121. [https://doi.org/10.1016/0010-0277\(91\)90033-Z](https://doi.org/10.1016/0010-0277(91)90033-Z).
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge: MIT Press.
- Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. Greenwood Publishing Group.
- Chondrogianni, Vasiliki & Hyowon Kwon. 2019. The development of English tense and agreement morphology in Welsh–English bilingual children with and without specific language impairment. *Applied Psycholinguistics* 1–32. <https://doi.org/10.1017/S0142716418000772>.
- Christensen, Peer, Riccardo Fusaroli & Kristian Tylén. 2016. Environmental constraints shaping constituent order in emerging communication systems: Structural iconicity, interactive alignment and conventionalization. *Cognition* 146. 67–80. <https://doi.org/10.1016/j.cognition.2015.09.004>.
- Christiansen, Morten H. & Nick Chater. 2008. Language as shaped by the brain. *Behavioral and Brain Sciences* 31(5). 489–509. <https://doi.org/10.1017/S0140525X08004998>.
- Christiansen, Morten H. & Nick Chater. 2016. The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences* 39. <https://doi.org/10.1017/S0140525X1500031X>.
- Christiansen, Morten H. & Simon Kirby (eds.). 2003. *Language Evolution*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199244843.001.0001>.
- Clark, Andy & David Chalmers. 1998. The Extended Mind. *Analysis*. [Analysis Committee, Oxford University Press] 58(1). 7–19.
- Clark, Herbert H. 1996. *Using Language*. Cambridge: Cambridge University Press.
- Clark, Herbert H. & Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition* 22(1). 1–39. [https://doi.org/10.1016/0010-0277\(86\)90010-7](https://doi.org/10.1016/0010-0277(86)90010-7).

- Cogswell, Michael, Jiasen Lu, Stefan Lee, Devi Parikh & Dhruv Batra. 2019. Emergence of Compositional Language with Deep Generational Transmission. <https://arxiv.org/abs/1904.09067v2> (28 October, 2020).
- Cohen, Jacob. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 20(1). 37–46. <https://doi.org/10.1177/001316446002000104>.
- Community, Blender Online. 2018. *Blender - a 3D modelling and rendering package*. Amsterdam: Stichting Blender Foundation. <http://www.blender.org>.
- Corballis, Michael C. 2003. From mouth to hand: Gesture, speech, and the evolution of right-handedness. *Behavioral and Brain Sciences* 26(02). 199–260. <https://doi.org/10.1017/S0140525X03000062>.
- Corballis, Michael C. 2012. The origins of language in manual gestures. In Maggie Tallerman & Kathleen R. Gibson (eds.), *The Oxford Handbook of Language Evolution*, 382–386. Oxford: Oxford University Press.
- 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. <https://doi.org/10.1126/sciadv.aaw2594>.
- Cremers, Jolien & Irene Klugkist. 2018. One Direction? A Tutorial for Circular Data Analysis Using R With Examples in Cognitive Psychology. *Frontiers in Psychology*. Frontiers 9. <https://doi.org/10.3389/fpsyg.2018.02040>.
- Croft, William. 2000. *Explaining Language Change: An Evolutionary Approach*. Harlow: Pearson Longman.
- Culbertson, Jennifer & David Adger. 2014. Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences* 111(16). 5842–5847. <https://doi.org/10.1073/pnas.1320525111>.
- Culbertson, Jennifer & Simon Kirby. 2016. Simplicity and Specificity in Language: Domain-General Biases Have Domain-Specific Effects. *Frontiers in Psychology* 6. <https://doi.org/10.3389/fpsyg.2015.01964>.
- Culbertson, Jennifer, Marieke Schouwstra & Simon Kirby. 2020. From the world to word order: Deriving biases in noun phrase order from statistical properties of the world. *Language*. Linguistic Society of America 96(3). 696–717. <https://doi.org/10.1353/lan.2020.0045>.
- Danziger, Eve. 2010. Deixis, gesture, and cognition in spatial Frame of Reference typology. *Studies in Language* 34(1). 167–185. <https://doi.org/10.1075/sl.34.1.16dan>.
- Danziger, Eve & Eric Pederson. 1998. Through the Looking Glass: Literacy, Writing Systems and Mirror-Image Discrimination. *Written Language & Literacy* 1(2). 153–169. <https://doi.org/10.1075/wll.1.2.02dan>.
- Darwin, Charles. 1871. *The Descent of Man*. London: Murray.
- Dasen, Pierre R. & Ramesh C. Mishra. 2010. *Development of geocentric spatial language and cognition: An eco-cultural perspective*. Vol. 12. Cambridge University Press.
- Deacon, Terrence William. 1998. *The symbolic species: the co-evolution of language and the brain*. 1st ed. New York: W.W. Norton.
- Dediu, Dan & Bart de Boer. 2016. Language evolution needs its own journal. *Journal of Language Evolution*. Oxford Academic 1(1). 1–6. <https://doi.org/10.1093/jole/lzv001>.
- Dediu, Dan, Michael Cysouw, Stephen C. Levinson, Andrea Baronchelli, Morten H. Christiansen, William Croft, Nicholas Evans, et al. 2013. Cultural evolution of language. In Peter J. Richerson & Morten H. Christiansen (eds.), *Cultural evolution: Society, technology, language, and religion*, 303–332. Cambridge, Mass: 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*

- (The Multimodal Origins of Linguistic Communication) 54. 9–20.
<https://doi.org/10.1016/j.langcom.2016.10.002>.
- Dediu, Dan & D. Robert Ladd. 2007. Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *Proceedings of the National Academy of Sciences* 104(26). 10944–10949. <https://doi.org/10.1073/pnas.0610848104>.
- Dediu, Dan & Stephen C. Levinson. 2013. On the antiquity of language: the reinterpretation of Neanderthal linguistic capacities and its consequences. *Frontiers in Psychology* 4. <https://doi.org/10.3389/fpsyg.2013.00397>.
- Dediu, Dan & Stephen C. Levinson. 2018. Neanderthal language revisited: not only us. *Current Opinion in Behavioral Sciences* 21. 49–55. <https://doi.org/10.1016/j.cobeha.2018.01.001>.
- Diessel, Holger. 2014. Demonstratives, frames of reference, and semantic universals of space. *Linguistics and Language Compass* 8(3). 116–132. <https://doi.org/10.1111/lnc3.12066>.
- Dingemanse, Mark. 2012. Advances in the Cross-Linguistic Study of Ideophones: Advances in the Cross-Linguistic Study of Ideophones. *Language and Linguistics Compass* 6(10). 654–672. <https://doi.org/10.1002/lnc3.361>.
- Dingemanse, Mark, Damián E. Blasi, Gary Lupyan, Morten H. Christiansen & Padraic Monaghan. 2015. Arbitrariness, Iconicity, and Systematicity in Language. *Trends in Cognitive Sciences* 19(10). 603–615. <https://doi.org/10.1016/j.tics.2015.07.013>.
- Dirksmeyer, Tyko. 2008. *Spatial deixis in Chintang: Aspects of a grammar of space*. Universität Leipzig Leipzig. <http://www.tyko.de/ccount/click.php?id=2> (29 October, 2020).
- Donelson, Katharine. 2018. *On Space in Conversation: Coordinating Perspectives in Spatial Reference*. State University of New York at Buffalo PhD Thesis.
- Dor, Danny, Chris Knight & Jerome Lewis (eds.). 2014. *The social origins of language* (Oxford Studies in the Evolution of Language 19). First edition. New York: Oxford University Press.
- Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. *WALS Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Dunbar, Robin. 1996. *Grooming, gossip, and the evolution of language*. Cambridge: Harvard University Press.
- Ebert, Karen H. 1999. The UP-DOWN dimension in Rai grammar and mythology. *Himalayan space: Cultural horizons and practices*. Völkerkunde Museum Zürich 107. 134.
- Eggleston, Alyson, Elena Benedicto & Mayangna Yulbarangyang Balna. 2011. Spatial frames of reference in Sumu-Mayangna. *Language Sciences* 33(6). 1047–1072.
- Ericsson, K Anders & Walter Kintsch. 1995. Long-term working memory. *Psychological review*. American Psychological Association 102(2). 211.
- Errico, Francesco d' & Marian Vanhaeren. 2009. Earliest personal ornaments and their significance for the origin of language debate. In Rudolf Botha & Chris Knight (eds.), *The cradle of language*, 16–40. Oxford: Oxford University Press.
- Evans, Nicholas. 2013. Language diversity as a tool for understanding cultural evolution. In Peter J. Richerson & Morten H. Christiansen (eds.), *Cultural Evolution: Society, Technology, Language, and Religion*, 233–268. MIT Press.
- Evans, Nicholas & Stephen C. Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32(5). 429–448. <https://doi.org/10.1017/S0140525X0999094X>.
- Everett, Caleb. 2013a. *Linguistic relativity: Evidence across languages and cognitive domains*. Vol. 25. Walter de Gruyter.
- Everett, Caleb. 2013b. Evidence for Direct Geographic Influences on Linguistic Sounds: The Case of Ejectives. *PLOS ONE* 8(6). e65275. <https://doi.org/10.1371/journal.pone.0065275>.

- Everett, Caleb. 2017. Languages in Drier Climates Use Fewer Vowels. *Frontiers in Psychology* 8. <https://doi.org/10.3389/fpsyg.2017.01285>.
- 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. <https://doi.org/10.1073/pnas.1417413112>.
- Everett, Caleb, Damián E. Blasi & Seán G. Roberts. 2016a. Response: Climate and language: has the discourse shifted? *Journal of Language Evolution* 1(1). 83–87. <https://doi.org/10.1093/jole/lzv013>.
- Everett, Caleb, Damián E. Blasi & Seán G. Roberts. 2016b. Language evolution and climate: the case of desiccation and tone. *Journal of Language Evolution*. Oxford Academic 1(1). 33–46. <https://doi.org/10.1093/jole/lzv004>.
- Fay, Nicolas, Michael Arbib & Simon Garrod. 2013. How to Bootstrap a Human Communication System. *Cognitive Science* 37(7). 1356–1367. <https://doi.org/10.1111/cogs.12048>.
- Fay, Nicolas & T. Mark Ellison. 2013. The Cultural Evolution of Human Communication Systems in Different Sized Populations: Usability Trumps Learnability. (Ed.) Christine A. Caldwell. *PLoS ONE* 8(8). e71781. <https://doi.org/10.1371/journal.pone.0071781>.
- Fay, Nicolas, T. Mark Ellison & Simon Garrod. 2014. Iconicity: From sign to system in human communication and language. *Pragmatics & Cognition* 22(2). 244–263. <https://doi.org/10.1075/pc.22.2.05fay>.
- Fay, Nicolas, Simon Garrod, Leo Roberts & Nik Swoboda. 2010. The Interactive Evolution of Human Communication Systems. *Cognitive Science* 34(3). 351–386. <https://doi.org/10.1111/j.1551-6709.2009.01090.x>.
- Fay, Nicolas, Casey J Lister, T. Mark Ellison & Susan Goldin-Meadow. 2014. Creating a communication system from scratch: Gesture beats vocalization hands down. *Frontiers in Psychology* 5(APR). 1–12. <https://doi.org/10.3389/fpsyg.2014.00354>.
- Fedden, Sebastian & Lera Boroditsky. 2012. Spatialization of time in Mian. *Frontiers in Psychology* 3(NOV). 1–9. <https://doi.org/10.3389/fpsyg.2012.00485>.
- Feldmann, Horst. 2019. Do Linguistic Structures Affect Human Capital? The Case of Pronoun Drop: Linguistic Structures and Human Capital. *Kyklos* 72(1). 29–54. <https://doi.org/10.1111/kykl.12190>.
- Fisher, Nicholas I. 1995. *Statistical analysis of circular data*. Cambridge University Press.
- Fitch, W. Tecumseh. 2010. *The Evolution of Language*. Cambridge: Cambridge University Press.
- Fitch, W. Tecumseh. 2017. Empirical approaches to the study of language evolution. *Psychonomic Bulletin & Review* 24(1). 3–33. <https://doi.org/10.3758/s13423-017-1236-5>.
- Fodor, Jerry A. 1975. *The Language of Thought*. Cambridge: Harvard University Press.
- Fodor, Jerry A. 1983. *The modularity of mind*. MIT press.
- François, Quesque & Yves Rossetti. 2020. What Do Theory-of-Mind Tasks Actually Measure? Theory and Practice. *Perspectives on Psychological Science*. <https://doi.org/10.1177/1745691619896607>.
- Frith, Uta & Frederique de Vignemont. 2005. Egocentrism, allocentrism, and Asperger syndrome. *Consciousness and Cognition (The Brain and Its Self)* 14(4). 719–738. <https://doi.org/10.1016/j.concog.2005.04.006>.
- Fusaroli, Riccardo & Kristian Tylén. 2012. Carving language for social coordination: A dynamical approach. *Interaction studies* 13(1). 103–124.
- Fusaroli, Riccardo & Kristian Tylén. 2016. Investigating Conversational Dynamics: Interactive Alignment, Interpersonal Synergy, and Collective Task Performance. *Cognitive Science* 40(1). 145–171. <https://doi.org/10.1111/cogs.12251>.

- Galantucci, Bruno. 2005. An Experimental Study of the Emergence of Human Communication Systems. *Cognitive Science* 29(5). 737–767. https://doi.org/10.1207/s15516709cog0000_34.
- Galantucci, Bruno. 2009. Experimental Semiotics: A New Approach for Studying Communication as a Form of Joint Action. *Topics in Cognitive Science* 1(2). 393–410. <https://doi.org/10.1111/j.1756-8765.2009.01027.x>.
- Galantucci, Bruno & Simon Garrod. 2010. Experimental Semiotics: A new approach for studying the emergence and the evolution of human communication. *Interaction Studies* 11(1). 1–13.
- Galantucci, Bruno & Simon Garrod. 2011. Experimental Semiotics: A Review. *Frontiers in Human Neuroscience* 5. <https://doi.org/10.3389/fnhum.2011.00011>.
- Galantucci, Bruno, Simon Garrod & Gareth Roberts. 2012. Experimental Semiotics. *Language and Linguistics Compass* 6(8). 477–493. <https://doi.org/10.1002/lnc3.351>.
- Galati, Gaspare, Elie Lobel, Giuseppe Vallar, Alain Berthoz, Luigi Pizzamiglio & Denis Le Bihan. 2000. The neural basis of egocentric and allocentric coding of space in humans: a functional magnetic resonance study. *Experimental brain research*. Springer 133(2). 156–164.
- Gallistel, C. R. 2002. Conception, perception and the control of action. *Trends in Cognitive Sciences* 6(12). 504. [https://doi.org/10.1016/S1364-6613\(02\)02033-8](https://doi.org/10.1016/S1364-6613(02)02033-8).
- Gamer, Matthias, Jim Lemon, Maintainer Matthias Gamer, A Robinson & W Kendall's. 2012. Package 'irr.' *Various coefficients of interrater reliability and agreement*.
- Garrod, Simon & Anthony Anderson. 1987. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition* 27(2). 181–218. [https://doi.org/10.1016/0010-0277\(87\)90018-7](https://doi.org/10.1016/0010-0277(87)90018-7).
- Garrod, Simon & Gwyneth Doherty. 1994. Conversation, co-ordination and convention: an empirical investigation of how groups establish linguistic conventions. *Cognition* 53(3). 181–215. [https://doi.org/10.1016/0010-0277\(94\)90048-5](https://doi.org/10.1016/0010-0277(94)90048-5).
- Garrod, Simon, Nicolas Fay, John Lee, Jon Oberlander & Tracy MacLeod. 2007. Foundations of Representation: Where Might Graphical Symbol Systems Come From? *Cognitive Science* 31(6). 961–987. <https://doi.org/10.1080/03640210701703659>.
- Garrod, Simon & Martin J Pickering. 2004. Why is conversation so easy? *Trends in cognitive sciences*. Elsevier 8(1). 8–11.
- Garrod, Simon & Martin J. Pickering. 2009. Joint Action, Interactive Alignment, and Dialog. *Topics in Cognitive Science* 1(2). 292–304. <https://doi.org/10.1111/j.1756-8765.2009.01020.x>.
- Gibson, Edward, Richard Futrell, Julian Jara-Ettinger, Kyle Mahowald, Leon Bergen, Sivalogeswaran Ratnasingam, Mitchell Gibson, Steven T. Piantadosi & Bevil R. Conway. 2017. Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences* 114(40). 10785–10790. <https://doi.org/10.1073/pnas.1619666114>.
- Gibson, James J. 1979. *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Gladwin, Thomas. 1970. *East is a big bird: navigation and logic on Puluwat atoll*. Cambridge, Mass: Harvard University Press.
- Goldin-Meadow, S., W. C. So, A. Özyürek & C. Mylander. 2008. The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences* 105(27). 9163–9168. <https://doi.org/10.1073/pnas.0710060105>.
- Gong, Tao, Hangxian Gao, Zhen Wang & Lan Shuai. 2019. Perceptual constraints on colours induce the universality of linguistic colour categorisation. *Scientific Reports*. Nature Publishing Group 9(1). 7719. <https://doi.org/10.1038/s41598-019-44202-6>.
- 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. *Am J Psychiatry* 4.

- Gray, Russell D., Alexei J. Drummond & Simon J. Greenhill. 2009. Language Phylogenies Reveal Expansion Pulses and Pauses in Pacific Settlement. *Science* 323(5913). 479–483. <https://doi.org/10.1126/science.1166858>.
- Griffiths, Thomas L. & Michael L. Kalish. 2007. Language Evolution by Iterated Learning With Bayesian Agents. *Cognitive Science* 31(3). 441–480. <https://doi.org/10.1080/15326900701326576>.
- Haegeman, Liliane (ed.). 1997. *Elements of Grammar: Handbook in Generative Syntax*. Dordrecht: Kluwer.
- Hammarström, Harald. 2016. Linguistic diversity and language evolution. *Journal of Language Evolution* 1(1). 19–29. <https://doi.org/10.1093/jole/lzw002>.
- Hammarström, Harald, Robert Forkel & Martin Haspelmath. 2019. Glottolog 4.1. Jena. <https://glottolog.org/>.
- Harnad, Stevan. 1990. The symbol grounding problem. *Physica D: Nonlinear Phenomena* 42(1). 335–346. [https://doi.org/10.1016/0167-2789\(90\)90087-6](https://doi.org/10.1016/0167-2789(90)90087-6).
- Hartmann, Stefan. 2020. Language change and language evolution: Cousins, siblings, twins? *Glottology* 0(0). <https://doi.org/10.1515/glot-2020-2003>.
- Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687. <https://doi.org/10.1353/lan.2010.0021>.
- Haspelmath, Martin. 2020. Human Linguisticity and the Building Blocks of Languages. *Frontiers in Psychology* 10. <https://doi.org/10.3389/fpsyg.2019.03056>.
- Haspelmath, Martin, Matthew S. Dryer, David Gil & Bernard Comrie. 2005. The world atlas of language structures online. <http://wals.info>.
- Haun, Daniel B. M., C. J. Rapold, J. Call, G. Janzen & S. C. Levinson. 2006. Cognitive cladistics and cultural override in Hominid spatial cognition. *Proceedings of the National Academy of Sciences* 103(46). 17568–17573. <https://doi.org/10.1073/pnas.0607999103>.
- 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. <https://doi.org/10.1016/j.cognition.2010.12.009>.
- Haun, Daniel B.M. & Christian J. Rapold. 2009. Variation in memory for body movements across cultures. *Current Biology* 19(23). R1068–R1069. <https://doi.org/10.1016/j.cub.2009.10.041>.
- Hauser, Marc D., Noam Chomsky & W. Tecumseh Fitch. 2002. The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science* 298(5598). 1569–1579. <https://doi.org/10.1126/science.298.5598.1569>.
- Haviland, John B. 1993. Anchoring, Iconicity, and Orientation in Guugu Yimithirr Pointing Gestures. *Journal of Linguistic Anthropology* 3(1). 3–45. <https://doi.org/10.1525/jlin.1993.3.1.3>.
- Haviland, John B. 2013. The emerging grammar of nouns in a first generation sign language: Specification, iconicity, and syntax. *Gesture*. John Benjamins 13(3). 309–353. <https://doi.org/10.1075/gest.13.3.04hav>.
- Healey, Patrick GT. 2008. Interactive Misalignment: The Role of Repair in the Development of Group Sub-languages. 29.
- Healey, Patrick GT & Gregory Mills. 2006. Participation, precedence and co-ordination in dialogue. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, vol. 320. Erlbaum Mahwah, NJ.
- Heine, Bernd & Tania Kuteva. 2007. *The Genesis of Grammar: A Reconstruction*. Oxford: Oxford University Press.
- Heine, Bernd & Tania Kuteva. 2012. Grammaticalization Theory as a Tool for Reconstructing Language Evolution. In Maggie Tallerman & Kathleen R. Gibson (eds.), *The Oxford Handbook of Language Evolution*, 511–527. Oxford: Oxford University Press.

- 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. <https://doi.org/10.3758/s13428-015-0688-7>.
- Hockett, Charles F. 1960. The Origin of Speech. *Scientific American* 203(3). 88–97. <https://doi.org/10.2307/24940617>.
- Hoffmann, Thomas. 2011. *Preposition placement in English: A usage-based approach*. Cambridge: Cambridge University Press.
- Holler, Judith & Stephen C. Levinson. 2019. Multimodal Language Processing in Human Communication. *Trends in Cognitive Sciences* 23(8). 639–652. <https://doi.org/10.1016/j.tics.2019.05.006>.
- Howard, Matt C. 2017. A meta-analysis and systematic literature review of virtual reality rehabilitation programs. *Computers in Human Behavior* 70. 317–327. <https://doi.org/10.1016/j.chb.2017.01.013>.
- Hua, Xia, Simon J. Greenhill, Marcel Cardillo, Hilde Schneemann & Lindell Bromham. 2019. The ecological drivers of variation in global language diversity. *Nature Communications* 10(1). 2047. <https://doi.org/10.1038/s41467-019-09842-2>.
- Humboldt, Wilhelm von. 1836. *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluß auf die geistige Entwicklung des Menschengeschlechts*. Berlin: Akademie der Wissenschaften.
- Hurford, James R. 1990. Nativist and Functional Explanations in Language Acquisition. In I.M Roca (ed.), *Logical Issues in Language Acquisition*, 85–136. Dordrecht: Foris.
- Hurford, James R., Michael Studdert-Kennedy & Chris Knight (eds.). 1998. *Approaches to the Evolution of Language*. Cambridge: Cambridge University Press.
- Hutchby, Ian & Robin Wooffitt. 2008. *Conversation Analysis*. Polity.
- Hutchins, Edwin. 1983. Understanding micronesian navigation. *Mental models*. Lawrence Erlbaum Hillsdale, NJ 191–225.
- Hutchins, Edwin. 1995. *Cognition in the Wild*. MIT press.
- Hutchins, Edwin & Geoffrey E Hinton. 1984. Why the Islands Move. *Perception*. SAGE Publications Ltd STM 13(5). 629–632. <https://doi.org/10.1068/p130629>.
- Imai, Mutsumi & Sotaro Kita. 2014. The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Phil. Trans. R. Soc. B* 369(1651). 20130298.
- Imai, Mutsumi, Sotaro Kita, Miho Nagumo & Hiroyuki Okada. 2008. Sound symbolism facilitates early verb learning. *Cognition* 109(1). 54–65.
- Jackendoff, Ray. 1996. The architecture of the linguistic-spatial interface. In Paul Bloom, Merrill F. Garrett, Lynn Nadel & Mary A. Peterson (eds.), *Language and space*, vol. 1, 30. Cambridge, MA: The MIT Press.
- Jackson, Joshua Conrad, Joseph Watts, Teague R. Henry, Johann-Mattis List, Robert Forkel, Peter J. Mucha, Simon J. Greenhill, Russell D. Gray & Kristen A. Lindquist. 2019. Emotion semantics show both cultural variation and universal structure. *Science* 366(6472). 1517–1522. <https://doi.org/10.1126/science.aaw8160>.
- Kanwal, Jasmeen, Kenny Smith, Jennifer Culbertson & Simon Kirby. 2017. Zipf's Law of Abbreviation and the Principle of Least Effort: Language users optimise a miniature lexicon for efficient communication. *Cognition* 165. 45–52. <https://doi.org/10.1016/j.cognition.2017.05.001>.
- Kemp, Charles & Terry Regier. 2012. Kinship Categories Across Languages Reflect General Communicative Principles. *Science* 336(6084). 1049–1054. <https://doi.org/10.1126/science.1218811>.
- Kendon, Adam. 2017. Reflections on the “Gesture-First” hypothesis of language origins. *Psychonomic Bulletin & Review* 24(1). 163–170. <https://doi.org/10.3758/s13423-016-1117-3>.
- Kim, Jaehyeon, Yongtae Kim & Jian Zhou. 2017. *Languages and Earnings Management*. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network.

- Kirby, Kathryn R., Russell D. Gray, Simon J. Greenhill, Fiona M. Jordan, Stephanie Gomes-Ng, Hans-Jörg Bibiko, Damián E. Blasi, et al. 2016. D-PLACE: A Global Database of Cultural, Linguistic and Environmental Diversity. (Ed.) Alex Mesoudi. *PLOS ONE* 11(7). e0158391. <https://doi.org/10.1371/journal.pone.0158391>.
- Kirby, Simon. 2011. Language is an adaptive system: the role of cultural evolution in the origins of structure. *The Oxford Handbook of Language Evolution*. <https://doi.org/10.1093/oxfordhb/9780199541119.013.0061>.
- Kirby, Simon. 2017. Culture and biology in the origins of linguistic structure. *Psychonomic Bulletin & Review* 24(1). 118–137. <https://doi.org/10.3758/s13423-016-1166-7>.
- Kirby, Simon, Hannah Cornish & Kenny Smith. 2008. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences* 105(31). 10681–10686. <https://doi.org/10.1073/pnas.0707835105>.
- Kirby, Simon, Mike Dowman & Thomas L. Griffiths. 2007. Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences* 104(12). 5241–5245. <https://doi.org/10.1073/pnas.0608222104>.
- Kirby, Simon, Tom Griffiths & Kenny Smith. 2014. Iterated learning and the evolution of language. *Current Opinion in Neurobiology* (SI: Communication and Language) 28. 108–114. <https://doi.org/10.1016/j.conb.2014.07.014>.
- Kirby, Simon & James R. Hurford. 2002. The emergence of linguistic structure: An overview of the iterated learning model. In Angelo Cangelosi & Domenico Parisi (eds.), *Simulating the evolution of language*, 121–147. London: Springer London. https://doi.org/10.1007/978-1-4471-0663-0_6.
- Kirby, Simon, Monica Tamariz, Hannah Cornish & Kenny Smith. 2015. Compression and communication in the cultural evolution of linguistic structure. *Cognition* 141. 87–102. <https://doi.org/10.1016/j.cognition.2015.03.016>.
- Klatzky, Roberta L. 1998. Allocentric and Egocentric Spatial Representations: Definitions, Distinctions, and Interconnections. In Christian Freksa, Christopher Habel & Karl F. Wender (eds.), *Spatial Cognition: An Interdisciplinary Approach to Representing and Processing Spatial Knowledge* (Lecture Notes in Computer Science), 1–17. Berlin, Heidelberg: Springer. https://doi.org/10.1007/3-540-69342-4_1.
- Koolen, Ruud, Martijn Goudbeek & Emiel Kraemer. 2013. The effect of scene variation on the redundant use of color in definite reference. *Cognitive Science* 37(2). 395–411. <https://doi.org/10.1111/cogs.12019>.
- Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82(1). 1–26. <https://doi.org/10.18637/jss.v082.i13>.
- Ladd, D. Robert, Seán G. Roberts & Dan Dediu. 2015. Correlational Studies in Typological and Historical Linguistics. *Annual Review of Linguistics* 1(1). 221–241. <https://doi.org/10.1146/annurev-linguist-030514-124819>.
- Lakoff, George & Mark Johnson. 2003. *Metaphors we live by*. Chicago: University of Chicago Press.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. Vol. 1: Theoretical Prerequisites*. Stanford: Stanford University Press.
- Lawton, Carol A. 2001. Gender and regional differences in spatial referents used in direction giving. *Sex Roles* 44(5–6). 321–337.
- Lazaridou, Angeliki, Alexander Peysakhovich & Marco Baroni. 2016. Multi-Agent Cooperation and the Emergence of (Natural) Language. <https://arxiv.org/abs/1612.07182v2>.
- Le Guen, Olivier. 2011. Speech and Gesture in Spatial Language and Cognition Among the Yucatec Mayas. *Cognitive Science* 35(5). 905–938. <https://doi.org/10.1111/j.1551-6709.2011.01183.x>.

- Levinson, Stephen C. 1996. Frames of reference and Molyneux's question: Crosslinguistic evidence. In *Language and space* (Language, Speech, and Communication), 109–169. Cambridge, MA, US: The MIT Press.
- Levinson, Stephen C. 1997. Language and Cognition: The Cognitive Consequences of Spatial Description in Guugu Yimithirr. *Journal of Linguistic Anthropology* 7(1). 98–131. <https://doi.org/10.1525/jlin.1997.7.1.98>.
- Levinson, Stephen C. 2003. *Space in Language and Cognition: Explorations in Cognitive Diversity*. Cambridge University Press.
- 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. [https://doi.org/10.1016/S0010-0277\(02\)00045-8](https://doi.org/10.1016/S0010-0277(02)00045-8).
- Levinson, Stephen C. & David P. Wilkins. 2006. *Grammars of Space: Explorations in Cognitive Diversity*. Cambridge University Press.
- Levinson, Stephen, Sarah Cutfield, Michael Dunn, Nick Enfield, Sergio Meira & David Wilkins. 2018. *Demonstratives in cross-linguistic perspective*. Vol. 14. Cambridge University Press.
- Lewis, David. 1969. *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.
- Li, Hengshan, Panagiotis Mavros, Arnout Sabbe, Tyler Thrash, Victor Schinazi & Christoph Hölscher. 2018. User judgments of perceived and traversed distances in a real shopping mall and a corresponding virtual model. In *Cognitive Processing*, vol. 19. Springer. <https://doi.org/10.1007/s10339-018-0884-3>.
- Li, Peggy & Linda Abarbanell. 2018. Competing perspectives on frames of reference in language and thought. *Cognition* 170. 9–24. <https://doi.org/10.1016/j.cognition.2017.09.005>.
- Li, Peggy, Linda Abarbanell, Lila Gleitman & Anna Papafragou. 2011. Spatial reasoning in Tenejapan Mayans. *Cognition* 120(1). 33–53. <https://doi.org/10.1016/j.cognition.2011.02.012>.
- Li, Peggy & Lila Gleitman. 2002. Turning the tables: language and spatial reasoning. *Cognition* 83(3). 265–294. [https://doi.org/10.1016/S0010-0277\(02\)00009-4](https://doi.org/10.1016/S0010-0277(02)00009-4).
- Lin, Yen-Ting. 2017. *Spatial language and cognition in bilingual minds: Taiwan as a test case*. State University of New York at Buffalo PhD Thesis.
- Lindsey, Delwin T. & Angela M. Brown. 2002. Color Naming and the Phototoxic Effects of Sunlight on the Eye. *Psychological Science* 13(6). 506–512. <https://doi.org/10.1111/1467-9280.00489>.
- Little, Hannah, Kerem Eryilmaz & Bart de Boer. 2017. Signal dimensionality and the emergence of combinatorial structure. *Cognition* 168. 1–15. <https://doi.org/10.1016/j.cognition.2017.06.011>.
- Little, Hannah, Marcus Perlman & Kerem Eryilmaz. 2017. Repeated interactions can lead to more iconic signals. In *the 39th Annual Conference of the Cognitive Science Society (CogSci 2017)*, 760–765. Cognitive Science Society.
- Lockwood, Gwilym, Mark Dingemanse & Peter Hagoort. 2016. Sound-symbolism boosts novel word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 42(8). 1274–1281. <https://doi.org/10.1037/xlm0000235>.
- Lucy, John A. 2016. Recent Advances in the Study of Linguistic Relativity in Historical Context: A Critical Assessment: Linguistic Relativity: A Critical Assessment. *Language Learning* 66(3). 487–515. <https://doi.org/10.1111/lang.12195>.
- Lum, Jonathon & Jonathan Schlossberg. 2014. The Virtual Atoll Task: A spatial language elicitation tool. In *The 45th Australian Linguistic Society Conference Proceedings–2014 edited by Mark Harvey*.
- Lupyan, Gary & Rick Dale. 2010. Language Structure Is Partly Determined by Social Structure. *PLOS ONE* 5(1). e8559. <https://doi.org/10.1371/journal.pone.0008559>.

- 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.
- MacNeilage, Peter F. 2008. *The origin of speech* (Studies in the Evolution of Language 10). Oxford: Oxford University Press.
- Macuch Silva, Vinicius, Judith Holler, Asli Ozyurek & Seán G. Roberts. 2020. Multimodality and the origin of a novel communication system in face-to-face interaction. *Royal Society Open Science* 7(1). <https://doi.org/10.1098/rsos.182056>.
- Maddieson, Ian & Christophe Coupé. 2015. Human spoken language diversity and the acoustic adaptation hypothesis. *The Journal of the Acoustical Society of America* 138(3). 1838–1838. <https://doi.org/10.1121/1.4933848>.
- 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.
- Majid, Asifa, Seán G. Roberts, Ludy Cilissen, Karen Emmorey, Brenda Nicodemus, Lucinda O’Grady, Bencie Woll, et al. 2018. Differential coding of perception in the world’s languages. *Proceedings of the National Academy of Sciences* 115(45). 11369–11376. <https://doi.org/10.1073/pnas.1720419115>.
- Martin, Alexander, Theeraporn Ratitamkul, Klaus Abels, David Adger & Jennifer Culbertson. 2019. Cross-linguistic evidence for cognitive universals in the noun phrase. *Linguistics Vanguard* 5(1). <https://doi.org/10.1515/lingvan-2018-0072>.
- Martins, Pedro Tiago & Cedric Boeckx. 2019. Language evolution and complexity considerations: The no half-Merge fallacy. *PLOS Biology: Public Library of Science* 17(11). e3000389. <https://doi.org/10.1371/journal.pbio.3000389>.
- Mathy, Fabien & Jacob Feldman. 2012. What’s magic about magic numbers? Chunking and data compression in short-term memory. *Cognition* 122(3). 346–362. <https://doi.org/10.1016/j.cognition.2011.11.003>.
- McElreath, Richard. 2020. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan. 2nd Edition*. 1st edn. Chapman and Hall/CRC. <https://doi.org/10.1201/9781315372495>.
- McNeill, David. 2012. *How language began: Gesture and speech in human evolution*. Cambridge: Cambridge University Press.
- Meakins, Felicity, Caroline Jones & Cassandra Algy. 2016. Bilingualism, language shift and the corresponding expansion of spatial cognitive systems. *Language Sciences* 54. 1–13. <https://doi.org/10.1016/j.langsci.2015.06.002>.
- Meir, Irit, Mark Aronoff, Carl Börstell, So-One Hwang, Deniz Ilkbasaran, Itamar Kastner, Ryan Lopic, Adi Lifshitz Ben-Basat, Carol Padden & Wendy Sandler. 2017. The effect of being human and the basis of grammatical word order: Insights from novel communication systems and young sign languages. *Cognition* 158. 189–207. <https://doi.org/10.1016/j.cognition.2016.10.011>.
- Mendívil-Giró, José-Luis. 2019. Did language evolve through language change? On language change, language evolution and grammaticalization theory. *Glossa: a journal of general linguistics* 4(1). <https://doi.org/10.5334/gjgl.895>.
- Mesoudi, Alex & Alex Thornton. 2018. What is cumulative cultural evolution? *Proc. R. Soc. B* 285(1880). 20180712. <https://doi.org/10.1098/rspb.2018.0712>.
- Messerschmid, Ernst & Reinhold Bertrand. 2013. *Space stations: systems and utilization*. Springer Science & Business Media.
- Michel, J.-B., Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, The Google Books Team, J. P. Pickett, et al. 2011. Quantitative Analysis of Culture Using Millions of Digitized Books. *Science* 331(6014). 176–182. <https://doi.org/10.1126/science.1199644>.
- Miller, George A. & Philip N. Johnson-Laird. 1976. *Language and perception*. Belknap Press.

- 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. <https://doi.org/10.1080/00207590344000187>.
- Monaghan, Padraic, Nick Chater & Morten H. Christiansen. 2005. The differential role of phonological and distributional cues in grammatical categorisation. *Cognition* 96(2). 143–182. <https://doi.org/10.1016/j.cognition.2004.09.001>.
- Monaghan, Padraic, Morten H. Christiansen & Nick Chater. 2007. The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology* 55(4). 259–305. <https://doi.org/10.1016/j.cogpsych.2006.12.001>.
- Monaghan, Padraic, Karen Mattock & Peter Walker. 2012. The role of sound symbolism in language learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. US: American Psychological Association 38(5). 1152–1164. <https://doi.org/10.1037/a0027747>.
- Monaghan, Padraic, Richard C. Shillcock, Morten H. Christiansen & Simon Kirby. 2014. How arbitrary is language? *Philosophical Transactions of the Royal Society B: Biological Sciences*. Royal Society 369(1651). 20130299. <https://doi.org/10.1098/rstb.2013.0299>.
- Moore, Randi. 2018. *Spatial language and cognition in Isthmus Zapotec*. State University of New York at Buffalo PhD Thesis.
- Moran, Steven, Daniel McCloy & Richard Wright. 2014. PHOIBLE online. Max Planck Institute for Evolutionary Anthropology.
- Mordatch, Igor & Pieter Abbeel. 2017. Emergence of Grounded Compositional Language in Multi-Agent Populations. <https://arxiv.org/abs/1703.04908v2> (28 October, 2020).
- Morin, Olivier, Piers Kelly & James Winters. 2018. Writing, Graphic Codes, and Asynchronous Communication. *Topics in Cognitive Science*. <https://doi.org/10.1111/tops.12386>.
- Motamedi, Yasamin, Marieke Schouwstra, Kenny Smith, Jennifer Culbertson & Simon Kirby. 2019. Evolving artificial sign languages in the lab: From improvised gesture to systematic sign. *Cognition* 192. 103964. <https://doi.org/10.1016/j.cognition.2019.05.001>.
- Motamedi, Yasamin, Marieke Schouwstra, Kenny Smith & Simon Kirby. 2016. Linguistic Structure Emerges In The Cultural Evolution Of Artificial Sign Languages. In Seán G. Roberts, Christine Cuskley, Luke McCrohon, Lluís Barceló-Coblijn, Olga Fehér & Tessa Verhoeft (eds.), *The Evolution of Language. Proceedings of the 11th International Conference*.
- Müller, Cornelia, Alan J. Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill & Sedinha Tessoroff (eds.). 2013. *Body - language - communication: An international handbook on multimodality in human interaction* (Handbooks of Linguistics and Communication Science 38.1). Berlin, Boston: De Gruyter Mouton.
- Muller, Max. 1870. Darwinism tested by the science of language. Translated from the German of Professor August Schleicher. *Nature*. Nature Publishing Group 1(10). 256–259.
- Müller, Thomas F. & James Winters. 2018. Compression in cultural evolution: Homogeneity and structure in the emergence and evolution of a large-scale online collaborative art project. *PLOS ONE* 13(9). e0202019. <https://doi.org/10.1371/journal.pone.0202019>.
- Murray, Jeff W. 2017. *Building virtual reality with Unity and Steam VR*. CRC Press.
- Mylander, Carolyn & Susan Goldin-Meadow. 1991. Home sign systems in deaf children: The development of morphology without a conventional language model. *Theoretical issues in sign language research*. The University of Chicago Press Chicago, IL 2. 41–63.
- Naroll, Raoul. 1961. Two Solutions to Galton's Problem. *Philosophy of Science* 28(1). 15–39.
- Newcombe, Nora S. & Janelle Huttenlocher. 2000. *Making space: The development of spatial representation and reasoning*. MIT Press.
- Nielsen, Alan. 2016. *Systematicity, motivatedness, and the structure of the lexicon*. University of Edinburgh.

- Nölle, Jonas. 2015. *Environmental Constraints on Cultural Conventions: A Case Study of Linguistic Coordination in Joint Problem Solving*. Preprint. Thesis Commons. <https://doi.org/10.31237/osf.io/v8fr9>. <https://osf.io/v8fr9> (24 October, 2019).
- Nölle, Jonas, Riccardo Fusaroli, Gregory J. Mills & Kristian Tylén. 2020. Language as shaped by the environment: linguistic construal in a collaborative spatial task. *Palgrave Communications* 6(1). 1–10. <https://doi.org/10.1057/s41599-020-0404-9>.
- Nölle, Jonas, Stefan Hartmann & Peeter Tunits. 2020. Language evolution research in the year 2020: A survey of new directions. *Language Dynamics and Change*. Brill 10(1). 3–26. <https://doi.org/10.1163/22105832-bja10005>.
- Nölle, Jonas, Simon Kirby, Jennifer Culbertson & Kenny Smith. 2020. 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 (EvoLang13)*, 321–323.
- Nölle, Jonas & Michael Spranger. accepted. From the field into the lab: Causal approaches to the evolution of spatial language. *Linguistics Vanguard*.
- Nölle, Jonas, Marlene Staib, Riccardo Fusaroli & Kristian Tylén. 2018a. Supplementary information for “The emergence of systematicity: how environmental and communicative factors shape a novel communication system.” <https://osf.io/h5eas/>.
- Nölle, Jonas, Marlene Staib, Riccardo Fusaroli & Kristian Tylén. 2018b. The emergence of systematicity: How environmental and communicative factors shape a novel communication system. *Cognition* 181. 93–104. <https://doi.org/10.1016/j.cognition.2018.08.014>.
- O’Meara, Carolyn & Gabriela Pérez Báez. 2011. Spatial frames of reference in Mesoamerican languages. *Language Sciences* 33(6). 837–852. <https://doi.org/10.1016/j.langsci.2011.06.013>.
- Onuchic, José Nelson & Peter G Wolynes. 2004. Theory of protein folding. *Current Opinion in Structural Biology* 14(1). 70–75. <https://doi.org/10.1016/j.sbi.2004.01.009>.
- Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. *Science*. American Association for the Advancement of Science 349(6251). <https://doi.org/10.1126/science.aac4716>.
- Özçalışkan, Şeyda, Ché Lucero & Susan Goldin-Meadow. 2016. Does language shape silent gesture? *Cognition* 148. 10–18. <https://doi.org/10.1016/j.cognition.2015.12.001>.
- Pagel, Mark. 2009. Human language as a culturally transmitted replicator. *Nature Reviews Genetics*. Nature Publishing Group 10(6). 405–415. <https://doi.org/10.1038/nrg2560>.
- Pahunov, Denis. 2015. *Erosion Brush*. <https://assetstore.unity.com/packages/tools/terrain/erosion-brush-27389>.
- Palmer, Bill. 2002. Absolute spatial reference and the grammaticalisation of perceptually salient phenomena. In Giovanni Bennardo (ed.), *Representing space in Oceania: Culture in language and mind*, 107–157. Pacific Linguistics.
- Palmer, Bill. 2007. Pointing at the lagoon: Directional terms in Oceanic atoll-based languages. In Jeff Siegel, John Lynch & Diana Eades (eds.), *Language Description, History and Development*, 101–117. John Benjamins.
- Palmer, Bill. 2015. Topography in language: Absolute Frame of Reference and the Topographic Correspondence Hypothesis. In Rik De Busser & Randy J. LaPolla (eds.), *Cognitive Linguistic Studies in Cultural Contexts*, vol. 6, 177–226. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/clsc.6.08pal>. <https://benjamins.com/catalog/clsc.6.08pal> (21 March, 2018).
- 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. <https://doi.org/10.1515/lingty-2017-0011>.
- Pavlek, Barbara, James Winters & Olivier Morin. 2019. Ancient coin designs encoded increasing amounts of economic information over centuries. *Journal of Anthropological Archaeology* 56. 101103. <https://doi.org/10.1016/j.jaa.2019.101103>.
- Pearl, Judea. 2009. *Causality*. Cambridge university press.
- Pearl, Judea & Dana Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect*. 1st edn. USA: Basic Books, Inc.
- Pederson, Eric. 1993. Geographic and manipulable space in two Tamil linguistic systems. In *European conference on spatial information theory*, 294–311.
- Pederson, Eric, Eve Danziger, David Wilkins, Stephen C. Levinson, Sotaro Kita & Gunter Senft. 1998. Semantic Typology and Spatial Conceptualization. *Language* 74(3). 557–589. <https://doi.org/10.2307/417793>.
- Peeters, David. 2019. Virtual reality: A game-changing method for the language sciences. *Psychonomic Bulletin & Review*. <https://doi.org/10.3758/s13423-019-01571-3>.
- Peirce, Charles S. 1931. *Collected Papers of Charles Sanders Peirce*. (Ed.) Charles Hartshorne, Paul Weiss & Arthur W. Burks. Vol. 1. Harvard University Press Cambridge.
- Peirce, Jonathan W. 2007. PsychoPy—psychophysics software in Python. *Journal of neuroscience methods*. Elsevier 162(1–2). 8–13.
- Perlman, Marcus, Rick Dale & Gary Lupyan. 2015. Iconicity can ground the creation of vocal symbols. *Royal Society Open Science* 2(8). 150152. <https://doi.org/10.1098/rsos.150152>.
- Perlman, Marcus, Riccardo Fusaroli, Deborah Fein & Letitia Naigles. 2017. The use of iconic words in early child-parent interactions. In, 913–918. Cognitive Science Society.
- Perlman, Marcus & Gary Lupyan. 2018. People Can Create Iconic Vocalizations to Communicate Various Meanings to Naïve Listeners. *Scientific Reports*. Nature Publishing Group 8(1). 2634. <https://doi.org/10.1038/s41598-018-20961-6>.
- Perniss, Pamela. 2018. Why We Should Study Multimodal Language. *Frontiers in Psychology* 9. <https://doi.org/10.3389/fpsyg.2018.01109>.
- Perniss, Pamela, Robin Thompson & Gabriella Vigliocco. 2010. Iconicity as a General Property of Language: Evidence from Spoken and Signed Languages. *Frontiers in Psychology* 1. <https://doi.org/10.3389/fpsyg.2010.00227>.
- Perniss, Pamela & Gabriella Vigliocco. 2014. The bridge of iconicity: From a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1651). 20130300–20130300. <https://doi.org/10.1098/rstb.2013.0300>.
- Perry, Lynn K., Marcus Perlman & Gary Lupyan. 2015. Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PloS one* 10(9). e0137147.
- Petré, Peter & Freek Van de Velde. 2018. The real-time dynamics of the individual and the community in grammaticalization. *Language* 94(4). 867–901.
- Piaget, Jean & Bärbel Inhelder. 1956. *The Child's Conception of Space*. London: Routledge and Kegan Paul.
- Pickering, Martin J. & Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27(02). 169–226. <https://doi.org/10.1017/S0140525X04000056>.
- Pinker, Steven. 1994. *The Language Instinct: How the Mind Creates Language*. New York: Morrow.
- Pinker, Steven. 2007. *The stuff of thought: Language as a window into human nature*. Penguin.

- Pinker, Steven & Paul Bloom. 1990. Natural language and natural selection. *Behavioral and Brain Sciences* 13. 707–784.
- Pleyer, Michael, Stefan Hartmann, James Winters & Jordan Zlatev. 2017. Interaction and iconicity in the evolution of language: Introduction to the special issue. *Interaction Studies* 18(3). 303–313. <https://doi.org/10.1075/is.18.3.01ple>.
- Polian, Gilles & Jürgen Bohnemeyer. 2011. Uniformity and variation in Tzeltal reference frame use. *Language Sciences* 33(6). 868–891. <https://doi.org/10.1016/j.langsci.2011.06.010>.
- R Core Team. 2019. *R: A language and environment for statistical computing*. Vienna, Austria. <https://www.R-project.org/>.
- Rahimi, Kasra, Colin Banigan & Eric D. Ragan. 2020. Scene Transitions and Teleportation in Virtual Reality and the Implications for Spatial Awareness and Sickness. *IEEE Transactions on Visualization and Computer Graphics* 26(6). 2273–2287. <https://doi.org/10.1109/TVCG.2018.2884468>.
- 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. <https://doi.org/10.1098/rspb.2019.1262>.
- Regier, Terry, Alexandra Carstensen & Charles Kemp. 2016. Languages Support Efficient Communication about the Environment: Words for Snow Revisited. *PLOS ONE* 11(4). e0151138. <https://doi.org/10.1371/journal.pone.0151138>.
- Regier, Terry & P. Kay. 2004. Color Naming and Sunlight: Commentary on Lindsey and Brown (2002). *Psychological Science* 15(4). 289–290. <https://doi.org/10.1111/j.0956-7976.2004.00670.x>.
- Regier, Terry, Charles Kemp & Paul Kay. 2015. 11 word meanings across languages support efficient communication. *The handbook of language emergence*. Wiley Online Library 87. 237.
- Ren, Yi, Shangmin Guo, Serhii Havrylov, Shay B Cohen & Simon Kirby. 2019. Enhance the Compositionality of Emergent Language by Iterated Learning. In *3rd NeurIPS Workshop on Emergent Communication (EmeCom @ NeurIPS 2019)*, 10.
- Richerson, Peter J. & Robert Boyd. 2008. *Not by genes alone: How culture transformed human evolution*. University of Chicago press.
- Ritt, Nikolaus. 2004. *Selfish Sounds and Linguistic Evolution: A Darwinian Approach to Language Change*. Cambridge: Cambridge University Press.
- Roberts, Gareth. 2017. The linguist’s Drosophila: Experiments in language change. *Linguistics Vanguard* 3(1). <https://doi.org/10.1515/lingvan-2016-0086>.
- Roberts, Gareth & Maryia Fedzechkina. 2018. Social biases modulate the loss of redundant forms in the cultural evolution of language. *Cognition* 171. 194–201. <https://doi.org/10.1016/j.cognition.2017.11.005>.
- Roberts, Gareth, Jirka Lewandowski & Bruno Galantucci. 2015. How communication changes when we cannot mime the world: Experimental evidence for the effect of iconicity on combinatoriality. *Cognition* 141. 52–66. <https://doi.org/10.1016/j.cognition.2015.04.001>.
- Roberts, Seán G. 2018. Robust, Causal, and Incremental Approaches to Investigating Linguistic Adaptation. *Frontiers in Psychology* 9. <https://doi.org/10.3389/fpsyg.2018.00166>.
- Roberts, Seán G., Angarika Deb, Catherine Sheard, Simon J. Greenhill, Kaius Sinnemäki, José Segovia-Martín, Jonas Nölle, et al. 2020. CHIELD: The causal hypotheses in Evolutionary Linguistics Database. *Journal of Language Evolution* 5(1).
- Roberts, Seán G. & James Winters. 2012. Social Structure and Language Structure: the New Nominative Approach. *Psychology of Language and Communication*. Sciendo 16(2). 89–112. <https://doi.org/10.2478/v10057-012-0008-6>.
- Roberts, Seán G. & James Winters. 2013. Linguistic diversity and traffic accidents: Lessons from statistical studies of cultural traits. *PLoS one* 8(8). e70902.

- Roberts, Seán G., James Winters & Keith Chen. 2015. Future Tense and Economic Decisions: Controlling for Cultural Evolution. *PLOS ONE*. Public Library of Science 10(7). e0132145. <https://doi.org/10.1371/journal.pone.0132145>.
- Saussure, Ferdinand de. 1959. *Course in general linguistics*. New York: Philosophical Library.
- Schleicher, August. 1863. *Die Darwinische Theorie und die Sprachwissenschaft: Offenes Sendschreiben an Herrn Dr. Ernst Haeckel, o. Professor der Zoologie und Direktor des zoologischen Museums an der Universität Jena*. 2nd edn. Weimar: Böhlau.
- Schlossberg, Jonathan. 2018. *Atolls, islands and endless suburbia: Space and landscape in Marshallese*. University of Newcastle, Newcastle Australia.
- Schouwstra, Marieke & Henriëtte de Swart. 2014. The semantic origins of word order. *Cognition* 131(3). 431–436. <https://doi.org/10.1016/j.cognition.2014.03.004>.
- Schouwstra, Marieke, Henriëtte de Swart & Bill Thompson. 2019. Interpreting Silent Gesture: Cognitive Biases and Rational Inference in Emerging Language Systems. *Cognitive Science* 43(7). e12732. <https://doi.org/10.1111/cogs.12732>.
- Schück, Albert. 1882. Die astronomischen, geographischen und nautischen kenntnisse der Derwohoner der Karolinen und Marshall Inseln im Westlichen Grossen Ozean. *Aus Allen Welttheilen* 13. 51–57.
- Scott-Phillips, Thomas C., Simon Kirby & Graham R.S. Ritchie. 2009. Signalling signalhood and the emergence of communication. *Cognition* 113(2). 226–233. <https://doi.org/10.1016/j.cognition.2009.08.009>.
- Selten, Reinhard & Massimo Warglien. 2007. The emergence of simple languages in an experimental coordination game. *Proceedings of the National Academy of Sciences* 104(18). 7361–7366. <https://doi.org/10.1073/pnas.0702077104>.
- Senghas, A. 2004. Children Creating Core Properties of Language: Evidence from an Emerging Sign Language in Nicaragua. *Science* 305(5691). 1779–1782. <https://doi.org/10.1126/science.1100199>.
- Shapero, Joshua A. 2017. Does Environmental Experience Shape Spatial Cognition? Frames of Reference Among Ancash Quechua Speakers (Peru). *Cognitive Science* 41(5). 1274–1298. <https://doi.org/10.1111/cogs.12458>.
- Shusterman, Anna & Peggy Li. 2016. Frames of reference in spatial language acquisition. *Cognitive Psychology* 88. 115–161. <https://doi.org/10.1016/j.cogpsych.2016.06.001>.
- Silvey, Catriona, Simon Kirby & Kenny Smith. 2015. Word Meanings Evolve to Selectively Preserve Distinctions on Salient Dimensions. *Cognitive Science* 39(1). 212–226. <https://doi.org/10.1111/cogs.12150>.
- Slonimska, Anita & Seán G. Roberts. 2017. A case for systematic sound symbolism in pragmatics: Universals in wh-words. *Journal of Pragmatics* 116. 1–20. <https://doi.org/10.1016/j.pragma.2017.04.004>.
- Smith, Kenny, Simon Kirby & Henry Brighton. 2003. Iterated Learning: A Framework for the Emergence of Language. *Artificial Life* 9(4). 371–386. <https://doi.org/10.1162/106454603322694825>.
- Smith, Kenny, Monica Tamariz & Simon Kirby. 2013. Linguistic Structure is an Evolutionary Trade-Off between Simplicity and Expressivity. In Markus Knauff, Michael Pauen, Natalie Sebanz & Ipke Wachsmuth (eds.), *Cooperative Minds: Social Interaction and Group Dynamics*, 1348–1353. Austin, TX: Cognitive Science Society.
- Smith, Kenny & Elizabeth Wonnacott. 2010. Eliminating unpredictable variation through iterated learning. *Cognition* 116(3). 444–449. <https://doi.org/10.1016/j.cognition.2010.06.004>.

- Sneller, Betsy & Gareth Roberts. 2018. Why some behaviors spread while others don't: A laboratory simulation of dialect contact. *Cognition* 170. 298–311. <https://doi.org/10.1016/j.cognition.2017.10.014>.
- Spranger, Michael. 2011. Recruitment, Selection and Alignment of Spatial Language Strategies. In T. Lenaerts, M. Giacobini, H. Bersini, P. Bourguine, M. Dorigo & R. Doursat (eds.), vol. Advances in Artificial Life, ECAL 2011: Proceedings of the Eleventh European Conference on the Synthesis and Simulation of Living Systems, 771–778. MIT Press.
- Spranger, Michael. 2012a. The co-evolution of basic spatial terms and categories. In Steels Luc (ed.), *Experiments in cultural language evolution*, vol. 3, 111. Amsterdam: John Benjamins.
- Spranger, Michael. 2012b. Potential Stages in the Cultural Evolution of Spatial Language. In *The Evolution of Language*, 328–335. Kyoto, Japan: WORLD SCIENTIFIC. https://doi.org/10.1142/9789814401500_0043.
- Spranger, Michael. 2013. Evolving Grounded Spatial Language Strategies. *KI - Künstliche Intelligenz* 27(2). 97–106. <https://doi.org/10.1007/s13218-013-0245-4>.
- Spranger, Michael. 2016. *The evolution of grounded spatial language*. 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 (EVO LANG IX)*, 130–139.
- Steels, Luc. 1995. A self-organizing spatial vocabulary. *Artificial Life* 2(3). 319–332.
- Steels, Luc. 2000. Language as a Complex Adaptive System. In *Parallel Problem Solving from Nature PPSN VI* (Lecture Notes in Computer Science), 17–26. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-45356-3_2. https://link.springer.com/chapter/10.1007/3-540-45356-3_2 (27 June, 2018).
- Steels, Luc. 2008. The symbol grounding problem has been solved. so what's next. In Manuel de Vega, Arthur Glenberg & Arthur Graesser (eds.), *Symbols and embodiment: Debates on meaning and cognition*, 223–244. Oxford University Press.
- Steels, Luc. 2011. Modeling the cultural evolution of language. *Physics of Life Reviews* 8(4). 339–356. <https://doi.org/10.1016/j.plrev.2011.10.014>.
- Steinert-Threlkeld, Shane & Jakub Szymanik. 2019. Learnability and semantic universals. *Semantics and Pragmatics* 12(0). 4. <https://doi.org/10.3765/sp.12.4>.
- Stevens, Jon S. & Gareth Roberts. 2019. *Noise, economy, and the emergence of information structure in a laboratory language*. *Cognitive Science* 43(2). e12717. doi: 10.
- Sulik, J. & G. Lupyan. 2020. Appeals to “theory of mind” no longer explain much in language evolution. In A. Ravignani, C. Barbieri, M. Martins, M. Flaherty, Y. Jadoul, E. Lattenkamp, H. Little, K. Mudd & T. Verhoef (eds.), *The evolution of language: Proceedings of the 13th international conference (EvoLang13)*. <https://doi.org/10.17617/2.3190925>.
- Syrjänen, Kaj, Terhi Honkola, Jyri Lehtinen, Antti Leino & Outi Vesakoski. 2016. Applying Population Genetic Approaches within Languages. *Language Dynamics and Change* 6(2). 235–283. <https://doi.org/10.1163/22105832-00602002>.
- Tallerman, Maggie & Kathleen Rita Gibson (eds.). 2012. *The Oxford handbook of language evolution* (Oxford Handbooks in Linguistics). Oxford ; New York: Oxford University Press.
- Talmy, Leonard. 1985. Lexicalisation patterns: semantic structure in lexical forms. In Timothy Shopen (ed.), *Grammatical categories and the lexicon*. (Language Typology and Syntactic Description, Vol. III). Cambridge University Press.
- Talmy, Leonard. 2000. *Toward a cognitive semantics* (Language, Speech, and Communication). Cambridge, Mass: MIT Press.

- Tamariz, Monica. 2017. Experimental Studies on the Cultural Evolution of Language. *Annual Review of Linguistics* 3(1). 389–407. <https://doi.org/10.1146/annurev-linguistics-011516-033807>.
- Tamariz, Mónica & Simon Kirby. 2015. Culture: Copying, Compression, and Conventionality. *Cognitive Science* 39(1). 171–183. <https://doi.org/10.1111/cogs.12144>.
- Tarr, Michael J. & William H. Warren. 2002. Virtual reality in behavioral neuroscience and beyond. *Nature Neuroscience* 5(11). 1089–1092. <https://doi.org/10.1038/nn948>.
- Tenbrink, Thora. 2018. Linguistic spatial reference systems across everyday domains: How people talk about space in sailing, dancing, and horse riding. In *Cognitive Processing*, vol. 19, S31–S31.
- Tenbrink, Thora & Frank Dylla. 2017. Sailing: Cognition, action, communication. *Journal of Spatial Information Science* 2017(15). 3–33.
- Terrill, Angela & Niclas Burenhult. 2008. Orientation as a strategy of spatial reference. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language."* John Benjamins 32(1). 93–136. <https://doi.org/10.1075/sl.32.1.05ter>.
- The Inkscape Team. 2003. *Inkscape*. <https://inkscape.org/>.
- Theisen, Carrie Ann, Jon Oberlander & Simon Kirby. 2010. Systematicity and arbitrariness in novel communication systems. *Interaction Studies* 11(1). 14–32. <https://doi.org/10.1075/is.11.1.08the>.
- Thiering, Martin. 2014. *Spatial Semiotics and Spatial Mental Models: Figure-Ground Asymmetries in Language*. Mouton de Gruyter.
- Tinbergen, N. 1963. On aims and methods of Ethology. *Zeitschrift für Tierpsychologie* 20(4). 410–433.
- Tinits, Peeter, Jonas Nölle & Stefan Hartmann. 2017. Usage context influences the evolution of over-specification in iterated learning. *Journal of Language Evolution* 2(2). 148–159. <https://doi.org/10.1093/jole/lzx011>.
- Tomasello, Michael. 1999. *The Cultural Origins of Human Cognition*. Cambridge: Harvard University Press.
- Tomasello, Michael. 2008. *Origins of Human Communication*. Cambridge: MIT Press.
- Tomasello, Michael. 2014. *A natural history of human thinking*. Cambridge, Massachusetts ; London, England: Harvard University Press.
- Torreira, Francisco, Sara Bögels & Stephen C. Levinson. 2015. Breathing for answering: the time course of response planning in conversation. *Frontiers in Psychology*. Frontiers 6. <https://doi.org/10.3389/fpsyg.2015.00284>.
- Tylén, Kristian, Riccardo Fusaroli, Peer F. Bundgaard & Svend Østergaard. 2013. Making sense together: A dynamical account of linguistic meaning-making. *Semiotica* 2013(194). <https://doi.org/10.1515/sem-2013-0021>.
- Tylén, Kristian, Riccardo Fusaroli, Sergio Rojo, Katrin Heimann, Nicolas Fay, Niels N. Johannsen, Felix Riede & Marlize Lombard. 2020. The evolution of early symbolic behavior in Homo sapiens. *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.1910880117>.
- Van Staden, Miriam, Melissa Bowerman & Mariet Verhelst. 2006. Some properties of spatial description in Dutch. In Stephen C. Levinson & David P. Wilkins (eds.), *Grammars of Space: Explorations in Cognitive Diversity*. Cambridge University Press.
- Verhoef, Tessa, Simon Kirby & Bart de Boer. 2016. Iconicity and the Emergence of Combinatorial Structure in Language. *Cognitive Science* 40(8). 1969–1994. <https://doi.org/10.1111/cogs.12326>.
- Verhoef, Tessa, Simon Kirby & Bart De Boer. 2014. Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics*. Elsevier 43. 57–68.

- Vicario, Federico. 2008. Verbi analitici e organizzazione dello spazio cognitivo. *I verbi sintagmatici in italiano e nelle varietà dialettali: stato dell'arte e prospettive di ricerca: atti delle giornate di studio Torino, 19-20, febbraio 2007*. 31.
- Von Uexküll, Jakob. 1934. A stroll through the worlds of animals and men: A picture book of invisible worlds. *Semiotica* 89(4). 319–391. <https://doi.org/10.1515/semi.1992.89.4.319>.
- Von Uexküll, Jakob. 1992. A stroll through the worlds of animals and men: A picture book of invisible worlds. *Semiotica*. Walter de Gruyter, Berlin/New York 89(4). 319–391.
- Vos, Connie de & Roland Pfau. 2015. Sign Language Typology: The Contribution of Rural Sign Languages. *Annual Review of Linguistics* 1(1). 265–288. <https://doi.org/10.1146/annurev-linguist-030514-124958>.
- Wacewicz, Sławomir & Przemysław Zywicki. 2017. The multimodal origins of linguistic communication. *Language & Communication* 54. 1–8. <https://doi.org/10.1016/j.langcom.2016.10.001>.
- Wagemans, Johan, James H. Elder, Michael Kubovy, Stephen E. Palmer, Mary A. Peterson, Manish Singh & Rüdiger von der Heydt. 2012. A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin* 138(6). 1172–1217. <https://doi.org/10.1037/a0029333>.
- Wagenmakers, Eric-Jan, Tom Lodewyckx, Himanshu Kuriyal & Raoul Grasman. 2010. Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology* 60(3). 158–189. <https://doi.org/10.1016/j.cogpsych.2009.12.001>.
- Wassmann, Jürg & Pierre R. Dasen. 1998. Balinese spatial orientation: some empirical evidence of moderate linguistic relativity. *Journal of the Royal Anthropological Institute* 689–711.
- Weinreich, William Uriel, Labov & Marvin L. Herzog. 1968.
- Wertheimer, Max. 1923. Untersuchungen zur Lehre von der Gestalt. II. *Psychologische Forschung* 4(1). 301–350. <https://doi.org/10.1007/BF00410640>.
- Winter, Bodo, Marcus Perlman & Asifa Majid. 2018. Vision dominates in perceptual language: English sensory vocabulary is optimized for usage. *Cognition* 179. 213–220. <https://doi.org/10.1016/j.cognition.2018.05.008>.
- Winter, Bodo, Marcus Perlman, Lynn K. Perry & Gary Lupyan. 2017. Which words are most iconic?: Iconicity in English sensory words. *Interaction Studies*. John Benjamins 18(3). 443–464. <https://doi.org/10.1075/is.18.3.07win>.
- Winters, James, Simon Kirby & Kenny Smith. 2015. Languages adapt to their contextual niche. *Language and Cognition* 7(3). 415–449. <https://doi.org/10.1017/langcog.2014.35>.
- Winters, James, Simon Kirby & Kenny Smith. 2018. Contextual predictability shapes signal autonomy. *Cognition* 176. 15–30. <https://doi.org/10.1016/j.cognition.2018.03.002>.
- Woensdregt, Marieke, Chris Cummins & Kenny Smith. forthc. A computational model of the cultural co-evolution of language and mindreading. *Synthese*.
- Xu, Yang, Emmy Liu & Terry Regier. 2020. Numeral Systems Across Languages Support Efficient Communication: From Approximate Numerosity to Recursion. *Open Mind*. MIT Press 4. 57–70. https://doi.org/10.1162/opmi_a_00034.
- Xu, Yang, Terry Regier & Barbara C. Malt. 2016. Historical Semantic Chaining and Efficient Communication: The Case of Container Names. *Cognitive Science* 40(8). 2081–2094. <https://doi.org/10.1111/cogs.12312>.
- Zaslavsky, Noga, Charles Kemp, Terry Regier & Naftali Tishby. 2018. Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences* 115(31). 7937–7942. <https://doi.org/10.1073/pnas.1800521115>.
- Zinsli, Paul. 1945. *Grund und Grat: die Bergwelt im Spiegel der schweizerdeutschen Alpenmundarten*. A. Francke.

- Zipf, George Kingsley. 1948. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Ravenio Books.
- Zlatev, Jordan, Sławomir Wacewicz, Przemysław Żywicznyński & Joost van de Weijer. 2017. Multi-modal-first or pantomime-first?: Communicating events through pantomime with and without vocalization. *Interaction Studies* 18(3). 465–488. <https://doi.org/10.1075/is.18.3.08zla>.
- Żywicznyński, Przemysław, Sławomir Wacewicz & Marta Sibierska. 2018. Defining Pantomime for Language Evolution Research. *Topoi. An International Review of Philosophy* 37(2). 307–318. <https://doi.org/10.1007/s11245-016-9425-9>.
2020. *The ASJP Database*.