



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.

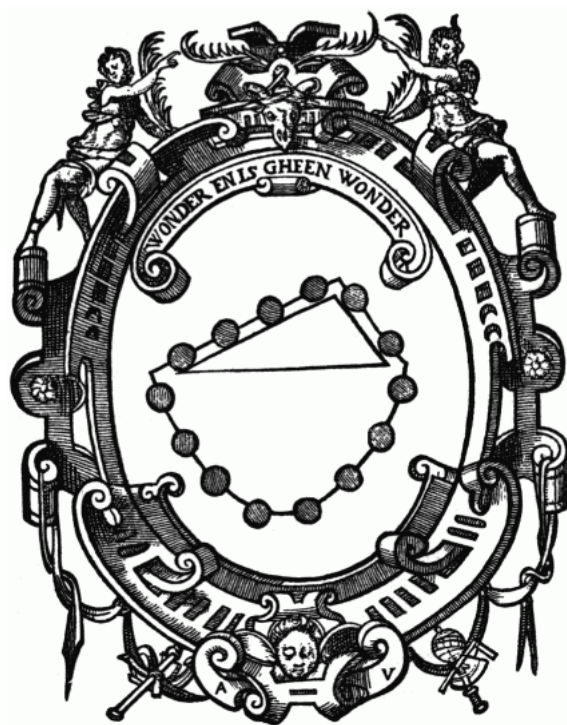
**Simplifying linguistic complexity:
culture and cognition in language evolution**

Carmen C. Saldana



Doctor of Philosophy
The School of Philosophy, Psychology and Language Sciences
The University of Edinburgh
2018

Wonder, not miracle .



Vignette from the cover of
De Beghinselen der Weeghconst (Stevin 1586).

Declaration

I declare that this thesis was composed by myself and the material contained therein has not been submitted for any other degree or qualification. The work reported is my own, except where explicitly stated otherwise in the text.

(Carmen C. Saldana)

Abstract

Languages are culturally transmitted through a repeated cycle of learning and communicative interaction. These two aspects of cultural transmission impose (at least) three interacting pressures that can shape the evolution of linguistic structure: a pressure for learnability, a pressure for expressivity, and a pressure for coordination amongst users in a linguistic community. This thesis considers how these sometimes competing pressures impact linguistic complexity across cultural time. Using artificial language and iterated learning experimental paradigms, I investigate the conditions under which complexity in morphological and syntactic systems emerges, spreads, and reduces. These experiments illustrate the interaction of transmission, learning and use in hitherto understudied domains—morphosyntax and word order.

In a first study (Chapter 2), I report the first iterated learning experiments to investigate the evolution of complexity in compositional structure at the word and sentence level. I demonstrate that a complex meaning space paired with pressures for learnability and communication can result in compositional hierarchical constituent structure, including fixed combinatorial rules of word formation and word order. This structure grants a productive and productively interpretable language and only requires learners to acquire a finite lexicon and a finite set of combinatorial rules (i.e., a grammar). In Chapter 3, I address the unique effect of communicative interaction on linguistic complexity, by removing language learning completely. Speakers use their native language to express novel meanings either in isolation or during communicative interaction. I demonstrate that even in this case, communicative interaction leads to more efficient and overall simpler linguistic systems.

These first two studies provide support for the claim that morphological and syntactic complexity are shaped by an overarching drive towards simplicity (or learnability) in language learning and communication. Chapter 4 reports a series of experiments assessing the possibility that the simplicity bias found in the first two studies operates at a different strength depending

on the linguistic level. Studies in natural language learning and in pidgin/creole genesis suggest that while morphological variation seems to be highly susceptible to regularisation, variation in other syntactic features, like word order, appears more likely to be reproduced. I test this experimentally by comparing regularisation of unconditioned variation across morphology and word order in the context of artificial language learning. I show that language users in fact regularise unconditioned variation in a similar way across linguistic levels, suggesting that the simplicity bias may be driven by a single, non-level-specific mechanism.

Taken together, the experimental evidence presented in this thesis supports the hypothesis that the cultural and cognitive pressures acting on language users during learning and communicative interaction—for learnability, expressivity and coordination—are at least partially responsible for the evolution of linguistic complexity. Specifically, they are responsible for the emergence of linguistic complexity which maximises learnability and communicative efficiency, and for the reduction of complexity which does not. More generally, the approach taken in this thesis promotes a view of complexity in linguistic systems as an evolving variable determined by the biases of language learners and users as languages are culturally transmitted.

Lay summary

Languages are transmitted over time through a repeated cycle of learning and communication. These two aspects of language transmission partially determine the way in which languages evolve. How well humans learn a language and how efficiently they can communicate with it determine how robustly the language is transmitted to future generations. A language is robustly transmitted over generations of learners if it is simple enough to be learned from the linguistic environment and effective enough to allow users to communicate successfully. At each generation, language users will help shift the language they receive towards one that better fits their learning and communicative needs. This drive towards simplicity affects the linguistic behaviour of language users, and the product of such behaviour accumulates over time to shape the way languages are organised.

This thesis is concerned with how language learning and use impact the evolution of complexity in linguistic systems. In particular, I investigate the conditions under which linguistic complexity emerges, spreads, and reduces. I investigate these questions using experiments in which participants are asked to learn and use artificial languages in a controlled environment that models crucial aspects of the learning and communicative contexts of natural languages.

In a first set of experiments I show how unstructured languages evolve linguistic structure as they are transmitted over generations of learners. At each generation, a participant has to learn and use an artificial language from the input produced by previous learners who acquired the language in the same way. The linguistic structure that evolves resembles the one we find in natural languages. Sentences are composed of words which at the same time are composed of different morphs (just as *cats* is composed of *cat* and *s*). In addition, the meaning of the sentences is determined not just by the individual meanings of its parts (the morphs), but also by the way in which those parts are combined (for instance, their position in the sentence). I

thus show that this type of complexity in linguistic structure evolves to facilitate learning and communication. In a second study I demonstrate that users describe meanings in their own language more simply and efficiently when they are required to communicate with another user than when they are not. The pressure to communicate successfully with a partner makes language users modify their behaviour to facilitate the communicative task. Finally, in a third set of experiments I investigate whether the drive towards simplicity to ease learning and communication impacts linguistic complexity in similar ways across different domains. I compare the extent to which users reduce redundant variability either in word order (syntax) or in word endings (morphology). I show that language users reduce such variation to similar degrees across linguistic domains, suggesting that a drive towards simplicity is ubiquitous in language.

Taken together, the experimental evidence presented in this thesis supports the hypothesis that learning and use in communication are at least partially responsible for the evolution of linguistic complexity. More generally, the approach taken in this thesis thus promotes a view of linguistic complexity as an evolving variable shaped by the pressures that act on the user during language learning and communication, and over time.

Acknowledgements

I have been tempted to allow my brother's voice to write the acknowledgements: "You are welcome", he proposed. However, I find it is not fair to just leave it to wit this time.

Firstly, I would like to thank the Obra Social "la Caixa" for the scholarship that funded my MSc and the first stages of my PhD. Thanks for the generous funding and thank you for reminding me that forces of resistance can be built even from within banks. Special thanks to Cedric Boeckx too, without him I would have never discovered evolutionary linguistics and would have never got the courage to apply for the scholarship to land on the Centre for Language Evolution (CLE).

Completing a PhD in my experience is a joint effort, in particular at the CLE and in particular surrounded by the families I am part of. I would like to start by thanking my (four!) supervisors for showing me week after week how language sciences move forward constructively. Thank you for sharing your professional struggles and victories with your students; you have provided me with the keys to build a productive, communicative and comfortable academic environment. Thank you very much for this legacy. Kenny, thank you for nodding along Greek and Latin words whilst allowing me to understand they are masks for simple concepts. Thank you also for letting me know when, by contrast, they were not masks. Hopefully this is not yet another one. Simon, I was once told you knew me as the "scary biolinguist" long before you knew my name; thank you for embracing my scariness as well as fear and help turn them into productivity and reflection when needed. I hope I can use what I learned from your support to now be there in your journey to name scary pseudonyms. Rob, I am very grateful to have you recruited as my supervisor after my first year review; you showed me that bridging the gap between theoretical and evolutionary linguistics is a complex but much needed task. Thank you very much for your thorough comments, and the constant support and believe in

my capacities. Jenny, although officially you have only supervised me for the last couple of months, I have benefited from your sharpness and humanity for three years now; thank you very much for the dialectic fights as much as for the support. I could not have possibly hoped for a better *super-vision* team.

I would like to continue by thanking my peer CLEers (and former LECers). It makes me very happy to know that I can thank you both professionally and personally. I thank my dearest “friend in history” Yasamin; we started an MSc five years ago, we struggled over regression models and bugs in code, and now we are pros! Thank you very much for making me always feel like we were in this together; we were always ready to turn around every time we looked over each other’s desk, and we will continue to. And *charaid* Jasmeen, what can I say to you?, we were also in this together very much indeed. We brought a whole new life and meaning to 5.01, and to writing up a thesis altogether: we made it fun (I can hear you think *yeah right* at this precise moment). I thank you now for all the (in)sanity and support, and I hope you take it as a voucher for the next 60 years. Special thanks to Jon too, and our long lasting discussions that always ended with us realising we agreed with each other probably from the beginning. I definitely hope you do not make me miss them; we have built *many* castles in the sky we some day need to inhabit and publish. I would also like to especially thank Mark and Matt, you were always available to help a newbie and even more so to share beers any time, anywhere: *muchas gracias a los dos*. I would like to also thank Kevin, my favourite developer-trained linguist; thanks to you I have been able to maintain several different Pythons in my computer for so long (I know your head might want to explode right now, but yes, you helped me tame such chaos).

I do not have space to thank individually every CLEer in the past four years, so this paragraph will be more name-heavy, but I mean every single one of them. Thank you Mónica for your heartening philosophy of life and science (and even politics), Olga for making the CLE a more interdisciplinary and fun place to work at, Crissy, Marieke Schouwstra, Vanessa Ferdinand, Bill Thompson, James Winters, Catriona Silvey, Sander Lestrade, Nikolaus Ritt, Alan Nielsen, Ashley Micklos and all the current PhD students at the CLE! I would also like to thank Toni and Katie for the best administrative support the CLE can hope for. Thanks you also to room 1.15 as a whole for creating such a friendly work environment. And last by not least, I would like to thank Mits, Josef, Pavel and Bert for providing such a great work environment as course organisers for tutors as well students; you have greatly contributed to making my

teaching experience as enriching as it has been.

If one does things correctly and procures a life outside work, other families are created. And if one is very lucky, families like mine are created. I want to thank my chosen family and comrades in Edinburgh: Alessio, Anna, Lucía, Ricardo, Pablo, Pol, Jasmeen, Minna and Panos. We helped each other make Edinburgh our home, and what a home we created! I also want to thank my chosen family back in Barcelona also known as Navidules: Mark, Gassol, Laia, Arche, and Anna. This would not have been possible without you, or better put, you would not have allowed it to. Thanks also to my two Sirs and fellow film critics, Edu and Miguel Ángel; to Ana and her unforgettable motivational speeches; to my fellow expatriate Clara; and to my family from *la facu*: Carola, Laura and Álvaro. I cannot wait to celebrate this feat with you all!

Not all families can be chosen, some of them are chosen for you, but I have also been very lucky with those ones and they have been a primary source of support throughout this journey. This is why I will never be able to express enough gratitude to my mother and to my brother and Alex in a page; I cannot write to you anything that I have not said already, and I am very proud of this. I am equally proud of knowing that Marieke, transversal to all these families, will understand why she is left to be thanked last. You support me in ways you nor I could verbalise, because this is your way, and I am very grateful for it. I promise you this in exchange: I will not make you proofread this page.

Contents

Abstract	viii
Lay Summary	ix
Acknowledgements	xi
1 Introduction	1
1.1 Introduction	1
1.2 Linguistic complexity	2
1.2.1 Background notions: local vs. global complexity and system vs. structural complexity	4
1.2.2 Relative complexity	8
1.2.3 Absolute complexity	11
1.2.4 Relationship between absolute and relative complexity	18
1.3 Linguistic complexity as an evolving variable	20
1.3.1 Communicative effectiveness and learnability in language transmission	22
1.3.2 Functionality in the evolution of linguistic complexity	27
1.4 Research focus and methodology	29
1.4.1 Methodological commitments	30
1.5 Outline of the thesis	32
2 The cultural evolution of complex compositional structure in the laboratory	35
2.1 Introduction	35
2.1.1 Structure in natural languages	36
2.1.2 The cultural evolution of linguistic structure	38

2.1.3	The evolution of complex linguistic structure: this study	40
2.2	Experiment 1: transmission and artificial pressure against ambiguity	43
2.2.1	Method	43
2.2.2	Measures	47
2.2.3	Analysis and results	51
2.2.4	Discussion	61
2.3	Experiment 2: transmission and communication	65
2.3.1	Method	65
2.3.2	Analyses and results	67
2.3.3	Discussion	76
2.4	Experiment 3: transmission, communication and artificial pressure against am- biguity	80
2.4.1	Method	80
2.4.2	Analyses and results	81
2.4.3	Discussion	88
2.5	Conclusion	92
2.6	Supplementary analyses	93
2.6.1	Learning error	93
2.6.2	Paradigmatic and syntagmatic complexity	94
2.6.3	Discussion	100
2.7	Summary	102
3	The effect of communicative interaction on efficiency and complexity	103
3.1	Introduction	103
3.2	Convention and communicative interaction	105
3.3	Least collaborative effort and linguistic complexity	107
3.4	A note on language learning and communicative interaction	110
3.5	Experiment 4	112
3.6	Materials and methods	114
3.6.1	Participants	114
3.6.2	Stimuli	114

3.6.3	Procedure	114
3.6.4	Experimental conditions	115
3.6.5	Measures	115
3.7	Analyses and results	120
3.7.1	Description length	120
3.7.2	Structural complexity	122
3.7.3	Lexical complexity	123
3.7.4	Lexical similarity	123
3.7.5	Expressivity	125
3.8	Discussion	126
3.8.1	The effect of communicative interaction on efficiency in production: description length over repeated use	126
3.8.2	The effect of communicative interaction on efficiency in comprehen- sion: lexical similarity	128
3.8.3	The effect of communicative interaction on linguistic complexity	129
3.9	Conclusion	130
4	The effect of level-specific linguistic variation on regularisation behaviour	133
4.1	Introduction	133
4.1.1	Level-specific effects on regularisation	135
4.2	The present study: regularisation behaviour across linguistic levels	138
4.3	Experiment 5: regularisation behaviour across linguistic levels in individuals	139
4.3.1	Materials and Methods	139
4.3.2	Analyses and results	144
4.3.3	Interim discussion	149
4.4	Experiment 6: the effect of the L1 two-modifier variant on regularisation be- haviour	151
4.4.1	Participants	151
4.4.2	Analyses and results	152
4.5	Discussion: the effect of linguistic level on regularisation in isolate production	156

4.6	Experiment 7: regularisation behaviour across linguistic levels in communicative interaction	158
4.6.1	Materials and methods	158
4.6.2	Analyses and results	159
4.6.3	Discussion: the effect of linguistic level on regularisation behaviour during communicative interaction	163
4.7	Encoding	163
4.7.1	Materials and methods	164
4.7.2	Results	165
4.8	General discussion	171
4.8.1	Regularisation behaviour is uniform across linguistic levels	171
4.8.2	Regularisation behaviour is comparable between isolate production and communicative interaction	172
4.8.3	Regularisation behaviour takes place in production as well as in learning	173
4.9	Conclusion	174
5	Summary and conclusions	177
5.1	Aims and contributions	177
5.1.1	Summary of experimental results	178
5.1.2	Roots of an overarching simplicity bias	182
5.2	Directions for future research	184
5.3	General conclusion	187
A	Relevant papers	217
A.1	Author contributions	217
A.2	Saldana, Smith, Kirby & Culbertson (2017) Is the strength of regularisation uniform across linguistic levels? <i>Proceedings of the 39th Annual Conference of the Cognitive Science Society</i> . Austin, TX: Cognitive Science Society.	217
B	Experimental data	225
B.1	Chapter 2	225
B.2	Chapter 3	225

B.3 Chapter 4	226
B.3.1 Experimental responses: languages and grammaticality judgments . . .	226
B.3.2 Post-experimental questionnaires	226

List of Figures

1.1	Taxonomy of complexity	3
2.1	Stimuli Kirby, Cornish, and Smith (2008)	41
2.2	Results Kirby et al. (2008)	41
2.3	Stimuli Experiments 1, 2 & 3	45
2.4	Form-meaning mappings Experiment 1	53
2.5	Structure Experiment 1	55
2.6	Syntactic nominal category Experiment 1	58
2.7	Word order types Experiment 1	59
2.8	Word order regularity Experiment 1	61
2.9	Homonymy Experiment 1	64
2.10	Form-meaning mappings Experiment 2	68
2.11	Homonymy Experiment 2	70
2.12	Communicative accuracy Experiment 2	71
2.13	Structure Experiment 2	72
2.14	Syntactic nominal category Experiment 2	74
2.15	Word order types Experiment 2	75
2.16	Word order regularity Experiment 2	76
2.17	Form-meaning mappings Experiment 3	82
2.18	Communicative accuracy Experiment 3	84
2.19	Structure Experiment 3	85
2.20	Syntactic nominal category Experiment 3	86
2.21	Word order types Experiment 3	89
2.22	Word order regularity Experiment 3	90

2.23 Homonymy Experiment 3	92
2.24 Learning error Experiments 1, 2 & 3	95
2.25 Paradigmatic complexity Experiments 1, 2 & 3	97
2.26 Syntagmatic complexity Experiments 1, 2 & 3	98
3.1 Example scene Experiment 4	120
3.2 Description length experiment 4	121
3.3 Lexical richness and similarity Experiment 4	124
4.1 Stimuli Experiments 5, 6 & 7	140
4.2 Entropy Experiment 5	145
4.3 Proportion majority input variant Experiment 5	147
4.4 Proportion two-modifier word orders Experiment 5	148
4.5 Proportion pre-nominal orders Experiment 5	150
4.6 Entropy Experiment 6	153
4.7 Proportion majority input variant Experiment 6	155
4.8 Proportion two-modifier word orders Experiment 6	155
4.9 Entropy Experiment 7	160
4.10 Misalignment Experiment 7	161
4.11 Proportion majority input variants in experimental and questionnaire responses: one-modifier variants	167
4.12 Proportion majority input variants in experimental and questionnaire responses: two-modifier variants	168

List of Tables

1.1	Examples of constituent order variation in French, Dutch and Korean.	6
1.2	Definite article systems in French, Dutch and German.	7
2.1	Nominals Experiment 1	54
2.2	Verbal elements Experiment 1	54
2.3	Constituent adjacency Experiment 1	57
2.4	Nominals Experiment 2	69
2.5	Verbal elements Experiment 2	69
2.6	Constituent adjacency Experiment 2	74
2.7	Nominals Experiment 3	83
2.8	Verbal elements Experiment 3	83
2.9	Constituent adjacency Experiment 3	86
3.1	Units lexical complexity Experiment 4	118
4.1	Probabilistic artificial language Experiment 5	141
4.2	Probabilistic artificial language Experiment 6	151
5.1	Simplified summary of the thesis' experimental results	179

Chapter 1

Introduction

The central task of a natural science is to make the wonderful commonplace: to show that complexity, correctly viewed, is only a mask for simplicity; to find pattern hidden in apparent chaos.

Simon (1996, p. 1)

1.1 Introduction

Languages are dynamic systems culturally transmitted over a repeated cycle of learning and communicative interaction. These two aspects of cultural transmission impose interacting selective pressures on the learner/user which, along with neutral evolutionary processes, shape the evolution of linguistic systems and their structure (Beckner et al. 2009; Christiansen & Chater 2008; Kirby & Hurford 2002; Mufwene 2013; K. Smith & Kirby 2012; Steels 2012). It is well established in the study of the cultural evolution of language that languages evolve (at least partly) over time to maximise their learnability as long as they do not jeopardise communicative effectiveness, or in other words, to minimise the effort of unambiguously conveying meaning (K. Smith & Kirby 2012).

What minimises effort in learning and communication? How does effort minimisation impact the complexity of the linguistic system and its structure? This thesis considers how the aforementioned pressures imposed on the learner/user impact *linguistic complexity* in mor-

phology and syntactic systems during language learning and communicative interaction, and over cultural time. Identifying the causal relationship between simplicity in learning and use and complexity in the linguistic system will ultimately help us understand the cognitive and cultural determinants that interact with non-linguistic and/or non-selective factors to shape not only complexity, but also simplification and complexification processes in language change and evolution. My contribution “to make the wonderful commonplace” in the language sciences will consist in uncovering patterns in the way individual cognition (language learning and use) and culture (social transmission) shape linguistic complexity.

In this chapter I will set out the theoretical framework from which I will address the research on the evolution of linguistic complexity in this thesis. In section 1.2 I will lay out a taxonomy of linguistic complexity encompassing the approaches, types and measures that will provide a background for evaluating the coverage of the studies in this thesis. In section 1.3 I will give an overview of previous work addressing language as a complex adaptive system, specially from studies exploring the cultural evolution of language; I will then characterise in more detail the pressures at play in language learning and communication and their interaction with complexity in linguistic systems and their structure. Section 1.4 will summarise the research focus of the thesis and methodological commitments. Finally, in Section 1.5, I will set out the outline of the thesis, mainly focusing on the content of the empirical chapters.

1.2 Linguistic complexity

Complexity is the property of a real world system that is manifest in the inability of any one formalism being adequate to capture all its properties. (Mikulecky 2001, p.344)

As highlighted by Mikulecky’s definition above, *complexity* has no universally accepted formalism in complexity science; *linguistic complexity* meets the same fate in the language sciences. A description of complexity is highly dependent on the viewpoint (e.g., language acquisition or linguistic typology); formalisms vary in the way they interact with a system and capture different aspects of complexity (e.g., the learnability of a system or the number of grammatical distinctions in a given linguistic domain). Since the turn of the century, the interest in linguistic complexity has increased notably; in the recent book *Complexity in language:*

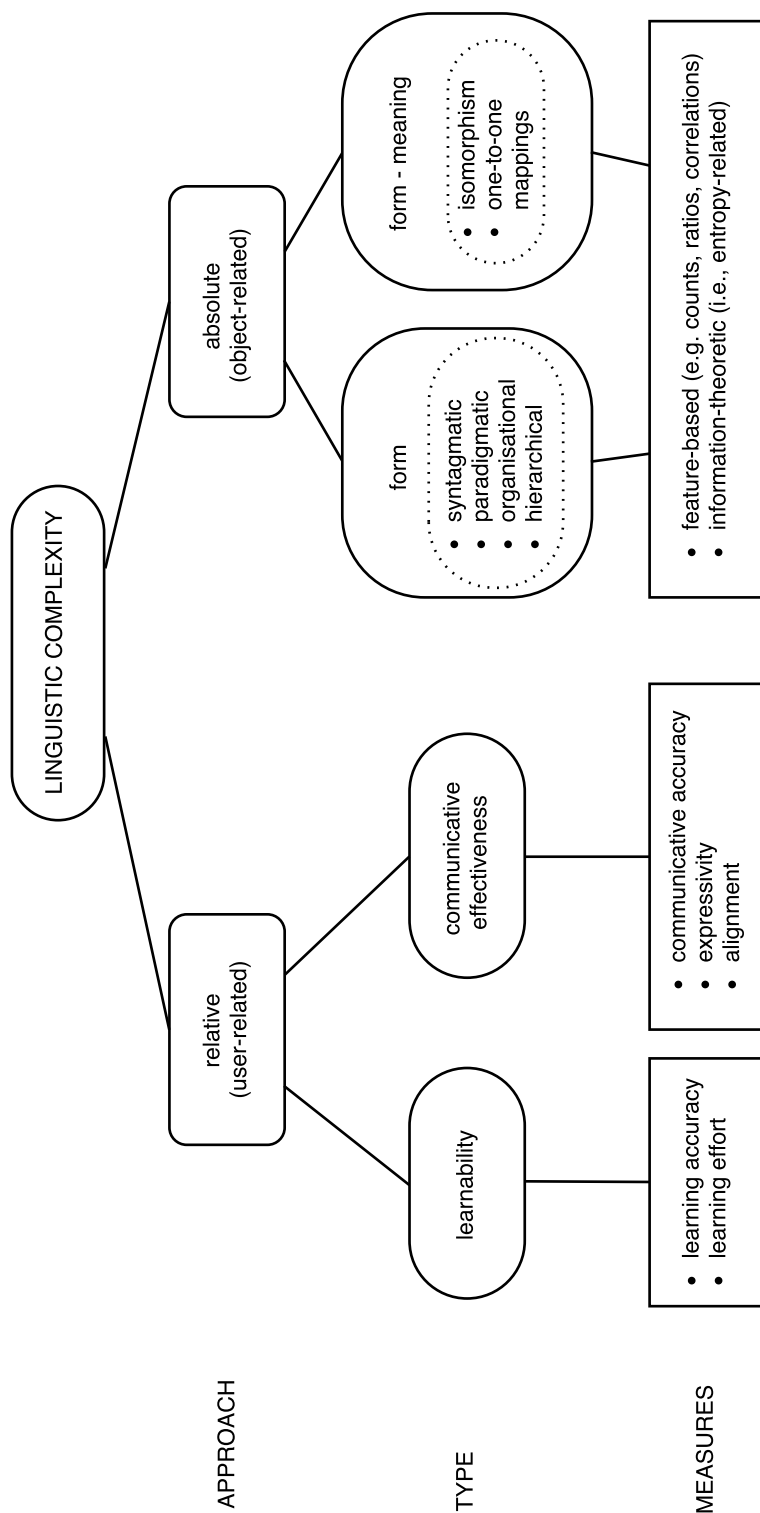


Figure 1.1: Taxonomy of approaches, types and measures of linguistic complexity comprised in this thesis.

developmental and evolutionary perspectives (Mufwene, Coupé, & Pellegrino 2017), the editors mention no less than 24 other books after 2001 (p. 1). Yet there is no consensus on the formulation of the notion of complexity in the field; probably as a symptom of the fundamental issue described by Mikulecky (for a list of measures used in complexity science, see Edmonds 1999; Lloyd 2001) and not of an inceptive discipline.

Notwithstanding, as in complexity sciences generally, the different formalisms and measures developed aim to answer the same questions about a system. In complexity sciences, these are three (Lloyd 2001; Page 2010; Rescher 1998): how hard is it to describe, how hard is it to create, and what is its degree of organisation (i.e., where does it lie between randomness and chaos). In the language sciences in particular, the questions linguists try to answer about a linguistic system or structure are: 1) how hard is it to acquire or process, and 2) how many and how variable are its component parts and/or their interactions. Question (1) approaches complexity in relation to the users' experience, and question (2) approaches it as a property of an autonomous object. Miestamo (2006b) coined the terms *relative* (user-related) and *absolute* (object-related) complexity to distinguish the focus of study in these two approaches, which rarely come together (cf. Mufwene et al. 2017). Thus, at a general level, linguistic complexity can be characterised in two ways: either by the cost/difficulty of their acquisition and/or processing in production or comprehension—relative complexity (e.g., Hawkins 2004, 2009; Kusters 2003, 2008; Szmrecsany & Kortmann 2009), or by “the number and variety of an item's constituent elements and the elaborateness of their interrelational structure” (Rescher 1998, p. 1)—absolute complexity (e.g., Dahl 2004; McWhorter 2005; Miestamo 2006b; Sinnemäki et al. 2011).

In order to characterise linguistic complexity in more detail, in the remainder of this section I discuss the taxonomy of approaches, types and measures of complexity developed for the purpose of this thesis. Figure 1.1 illustrates the taxonomy in question.

1.2.1 Background notions: local vs. global complexity and system vs. structural complexity

Local vs. global complexity A received view in linguistics considers that, although complexity might vary between domains, all languages will average to the same complexity:

Objective measurement is difficult, but impressionistically it would seem that the total grammatical complexity of any language, counting both morphology and syntax, is about the same as that for any other. This is not surprising, since all languages have about equally complex jobs to do. (Hockett 1960a, pp. 180–181)

[A] central finding of linguistics has been that all languages, both ancient and modern, spoken by both ‘primitive’ and ‘advanced’ societies, are equally complex in their structure. (Fortson IV 2004, pp. 180–181)

This view has led researchers in the field to establish a distinction between *local* and *global* complexity (Miestamo 2006b). Whereas local complexity is concerned with specific *parts* of the system (e.g., definite article paradigms or constituent orders, Figures 1.1 and 1.2 respectively), global complexity refers to the *overall* complexity of a language. This distinction has led linguists to conclude that whilst progress can be made in measuring local complexity across languages and their dynamics of change, measures of global complexity are unattainable (Deutscher 2009; Miestamo 2006b; Sinnemäki et al. 2011). Miestamo (2006b) identifies two general problems any attempt to measure global complexity faces: representativity and comparability. The former means that no metric can represent all aspects of a language. The problem of comparability, on the one hand, is concerned with the difficulty (or impossibility) of comparing different parts of languages in a meaningful way—e.g., how can we relate case marking and tone? On the other hand, it is about the impossibility of measuring the impact of each aspect of local complexity on global complexity. These issues raise a more fundamental question: why should a global metric be developed at all?

This thesis is concerned with how morphological and syntactic complexity evolves, and with how it is shaped by usage. I thus need to be able to assess different aspects of the language individually and not obscure changes in global complexity. Complexities are expected to vary locally within a language, as not all domains will interact in the same way with usage. I will thus exclusively measure local complexity and represent global complexity as the set of all local differences. Under this view, it is of no interest whether or not two languages are equally complex or not, but in which ways their complexity varies.

System vs. structural complexity Dahl (2004) established another important distinction

[French]

- (1) a. Rina donne un cadeau à Mina
 Rina gives a present to Mina
 ‘Rina gives a present to Mina’

[Dutch]

- (2) a. Rina geeft een cadeau aan Mina
 Rina gives a present to Mina
 ‘Rina gives a present to Mina’
- b. Rina geeft Mina een cadeau
 Rina gives Mina a present
 ‘Rina gives Mina a present’
- c. (dat) Rina een cadeau aan Mina geeft
 COMP Rina a present to Mina gives
 ‘(that) Rina gives a present to Mina’

[Korean]

- (3) a. Rina-ga Mina-ege seonmul-eul junda
 Rina-NOM Mina-DAT present-ACC gives
 ‘Rina gives a present to Mina’
- b. Rina-ga seonmul-eul Mina-ege junda
 Rina-NOM present-ACC Mina-DAT gives
 ‘Rina gives a present to Mina’
- c. Mina-ege seonmul-eul Rina-ga junda
 Mina-DAT present-ACC Rina-NOM gives
 ‘Rina gives a present to Mina’
- d. Mina-ga Rina-ege seonmul-eul junda
 Mina-NOM Rina-DAT present-ACC gives
 ‘Rina gives a present to Mina’
- e. seonmul-eul Rina-ga Mina-ege junda
 present-ACC Rina-NOM Mina-DAT gives
 ‘Rina gives a present to Mina’
- f. seonmul-eul Mina-ege Rina-ga junda
 present-ACC Mina-DAT Rina-NOM gives
 ‘Rina gives a present to Mina’
-

Table 1.1: Examples of constituent order variation in French, Dutch and Korean.

French		feminine	masculine	Dutch		common	neuter
	singular	la (or l')	le (or l')		singular	de	het
	plural	les	les		plural	de	de

German		feminine	masculine	neuter	plural
	Nominative	der	die	das	die
	Accusative	den	die	das	die
	Dative	dem	der	dem	den
Genitive	des	der	des	der	

Table 1.2: Definite article systems in French, Dutch and German.

between *system complexity*, a property of the language, and *structural complexity*¹, a property of individual expressions in a language. I will briefly illustrate the differences between these two concepts with examples of local absolute complexity.

Table 1.1 shows the constituent order variation (generally speaking, the order of subject, objects and verb) of three different languages, i.e., French, Dutch and Korean. For the sake of simplicity, the table only includes full forms and excludes pronominal forms as constituents. French has a fairly fixed SVO(O) constituent order (pronominal forms aside), as shown in 1a. Dutch, on the other hand, is a V2 language and alternates between verb-second and verb-final orders: whereas the matrix clause in 2a has the verb in second position, the embedded clause in 2c is verb-final. Moreover, sentences 2a-b show that Dutch (as English) allows dative alternation with some di-transitive verbs; however, it requires a change of form in the dative constituent, i.e., with or without *aan* ('to'). Unlike the other languages, Korean has case marking and although it has a canonical SO(O)V order, it allows the order of the subject and object to vary freely; sentences 3a-f show six different constituent orders without any modification of the constituent forms. If we consider order variability as an aspect of a system's complexity we can conclude that Korean is more complex than Dutch and Dutch is more complex than French. However, if we compare French and Dutch according to their definite article systems, Table 1.2 shows that French is more complex than Dutch as it has more distinctions between forms, and both are less complex than German. These comparisons pertain to whole morphological and syntactic systems and thus refer to *system complexity*.

By contrast, comparisons between patterns of individual expressions concern *structural complexity*. If we compare the dative alternation patterns in 2a-b in Dutch, we observe that 2a

¹Should not be confused with structural complexity in Rescher (1998).

requires an extra marker, i.e., the preposition *ann*. If we consider the number of elements in a sentence as an aspect of the sentence's complexity, we can conclude that 2a is more complex than 2b. And if we compare the number of morphs required to convey the meaning "Rina gives a present to Mina" across Dutch (2) and Korean (3), we can conclude that the structures in Korean—which include case-marking—are more complex than in Dutch. Structural complexity can also serve as the basis to measure complexity across a given sample of natural language expressions, e.g., by taking the average number of morphs per sentence.

Having established these background concepts, I will now proceed to discuss in detail the characterisation of complexity within relative (user-related) and absolute (object-related) approaches.

1.2.2 Relative complexity

In the relative or user-related approach, complexity is defined in relation to the experience of the language user; complexity is then measured in terms of cost or difficulty in acquisition and/or processing (Dahl 2004; Hawkins 2004; Kusters 2008; Miestamo 2006b). A relativist position thus always assumes a perspective or point of view from which linguistic complexity is evaluated; difficulty will always depend upon the interaction between aspects of language and the language user.

The main advantage of this approach is that it gives researchers the opportunity to straightforwardly evaluate linguistic complexity through the assessment of learners' and users' performances in naturalistic tasks and in psycholinguistic experiments. However, two main issues have been raised against the robustness of relative complexity, specially for cross-linguistic analyses. Firstly, difficulty in learning and processing is too dependent on the prior knowledge of the learner and on the context of use. Precisely its relativity makes it hard to achieve a consensual formulation of complexity as cost and/or difficulty: how should we decide which type of context of use and language user are primary for the definition of complexity? Secondly, and most problematically, there is no consensus on what is actually costly or difficult to the language learner/user either (Miestamo 2008).

Given the problems raised, several authors claim that cross-linguistic and typological studies should not rely on relative measures of complexity (Dahl 2004; Miestamo 2006b). However, discarding relative complexity altogether would be throwing the baby out with the bathwater.

Users are indeed constantly adapting their language to the context of use and thus from a process point of view of typology and for the study of language evolution more broadly, it is impossible to dismiss altogether: only with a good understanding of user-related complexity can we shed light on externally-motivated language change and variation. Whilst learning and communicative needs can vary between users and contexts, some might be shared: evolutionary linguistics has productively studied language emergence and change as a cultural product of pressures imposed onto the learner and user during language learning and use in communicative interaction (Boyd & Richerson 1988; Brighton, Smith, & Kirby 2005; Christiansen & Chater 2008; Kirby 1999; K. Smith et al. 2017; Steels 2012). The study of relative complexity in language evolution can benefit both from the study of general tendencies as well as individual variation, what I call *user-general* and *user-specific* relative complexity. Whereas user-specific complexity is relative to the specific mind and knowledge of an individual user, user-general complexity is relative to the human mind and knowledge—shared across language users. Additionally, grouping relative complexity into these two separate categories and acknowledging their differences will help elucidate the limits of the relativist approach in cross-linguistic generalisations.

User-specific complexity heavily depends on the user's specific experience and knowledge, linguistic or contextual. This particularly plays a role in the context of adult learning. Difficulty of acquisition of novel linguistic features will partially depend on the languages a learner knows and the contexts of exposure. Second language learners need to integrate novel grammars with their existing linguistic knowledge; thus at least initially, learners exploit prior knowledge, which can then interfere with the learning of the novel language (Ellis 2013; Odlin 1989; Weber, Christiansen, Petersson, Indefrey, & Hagoort 2016). To the extent that features overlap between languages, prior knowledge can serve as the basis for the novel feature and ease their acquisition; however, because similar features might differ in detail, the acquisition of the novel feature might also be hindered by prior knowledge (Ellis 2013; Odlin 1989). For instance, native speakers of Dutch (a V2 language), Korean (canonical SOV order) and French (fixed SVO order) might not find the learning of constituent order in German (V2) as complex: Dutch speakers will more likely match the target V2 order very early on whilst Korean and French speakers will start producing SOV and SVO orders initially and only later on match the target V2 (see Vainikka & Young-Scholten 1996). Difficulty of acquisition will also depend on the

context in which a learner receives data; e.g., learners who have previously learned a language in an academic environment might find it easier to acquire a novel language under similar circumstances and to perform similar tasks.

Given the variability in user-specific complexity, it does not come as a surprise that cross-linguistic analyses often dismiss it (Dahl 2011; McWhorter 2005; Miestamo 2006b). However, understanding this type of complexity and comparing it across languages and contexts is crucial to the identification of *user-general* relative complexity, which is of great importance for generalisations on externally-induced language change; without the former, we cannot test the latter. Experimentalists need to take into account that it is impossible to get rid of the influence of prior knowledge in an individual's performance. However, it is possible to identify common patterns after controlling—as thoroughly as possible—for variation in prior linguistic knowledge.

In the context of foreign language acquisition exclusively, Kusters (2003, 2008) has used the concept of the “generalised outsider” to discuss relative complexity: “[t]his person speaks a first language, and is not familiar with the second language in question, nor with the customs and background knowledge of the speech community” (p. 9). The notion of “generalised” in his case discards the consideration of interferences from prior linguistic knowledge from complexity assessments. In the context of language evolution, researchers also use the concept of a general learner/user. However, this is not explicitly restricted to the context of foreign language acquisition, or adult learning—although most experimental studies are looking at artificial language learning in adults (e.g. Kirby et al. 2008; Regier, Kemp, & Kay 2015). Across human cultures, languages are learned and used for communication. Linguistic complexity could then be estimated based on the efforts required by a “general” mind to acquire a language and successfully use it for communication. As noted in Mufwene (2008), “[a] linguistic system that fully responds with its structures to the communicative needs and the context of use may be seen of little complexity, while a complex linguistic system may be characteri[s]ed by structures that do not reflect these needs”. We can expand Mufwene's consideration to comprise learning as well as communication and propose that *features that in general terms balance the learnability of a system and its communicative effectiveness—in a given context—might be considered to be less complex*. Under these terms, systems are simple in relation to the user if they are *efficient*, i.e., minimise the effort required in language learning and production without

jeopardising effective communication (Fedzechkina, Jaeger, & Newport 2012; Kirby, Tamariz, Cornish, & Smith 2015; Regier et al. 2015; Zipf 1949).

How can we assess learnability and communicative effectiveness? Learnability can be measured by the accuracy and ease with which a learner reproduces the target grammar (Fehér, Wonnacott, & Smith 2016; Kirby et al. 2008; Michel 2011); given comparable amounts of input and in similar contexts, languages that can be reproduced with less effort by learners will be considered simpler. Communicative effectiveness can be measured by communicative accuracy or by the amount of information transmitted. If communication between interlocutors is successful, we can infer that it meets the requirements to clearly transmit information in a given context. Communicative success depends upon a shared system of conventions between interlocutors (i.e., alignment) and on the expressivity of the message to be conveyed in a given context; an ambiguous message will lower the chance of being interpreted correctly. The easier a message is to interpret, the simpler the system can be considered from the receiver's point of view. Effective communication thus entails that the message is expressive *enough* in a given context—full expressivity is unattainable (Levinson 2000). Effectiveness interacts with learning effort as well as production effort in natural language, i.e., with efficiency more broadly; a speaker/signer aims to spend the minimum time and resources required for unambiguously conveying a message to the receiver (H. H. Clark 1996; Zipf 1949). An utterance that conveys a meaning in context with less linguistic material will be considered simpler under these terms (for an alternative account, cf. Bisang 2009). Altogether, linguistic systems or structures that minimise the effort of unambiguously conveying a meaning will be considered to be efficient as they will maximise their learning and/or ease of production without jeopardising communicative effectiveness. Therefore, in terms of relative complexity and thus of cost/difficulty, the more efficiently a language satisfies learning and communicative needs, the simpler it is.

1.2.3 Absolute complexity

In the absolute or object-related approach, the definition of complexity is not related to the experiences of the language user but ascribed to the architecture of a system or structure (Dahl 2004). Broadly speaking, absolute complexity can be defined in terms of the number of parts of a system or structure, e.g., number of units, number of categories, number of rules or number of interactions between components. The more parts a system has, the more complex it is.

What constitutes a part is heavily theory-dependent and cross-linguistic analysis can only be taken as robust under comparable linguistic description (Haspelmath 2015).

There are two main questions the study of absolute complexity treats; how to measure the complexity of the form, and how to measure the complexity of the mapping between meaning and form (e.g., Kusters 2003; McWhorter 2005; Miestamo 2006b; Sinnemäki et al. 2011). For the purpose of this thesis, the form will be restricted to morphology and syntax. Within form and form-meaning mappings, there are different aspects which can vary in complexity. It is thus necessary to break down complexity into further categories. Based on previous work (Dahl 2004; Kusters 2008; McWhorter 2007; Miestamo 2006b; Rescher 1998), I differentiate the following types of absolute complexity for the purpose of this thesis²:

A. Complexity of the form

System complexity

- a. *Paradigmatic complexity*: the number of different types of components (e.g., grammatical distinctions), for example, the variety of definite articles (see Table 1.2). The more types in the paradigm, the higher the system complexity. For instance, the paradigm of definite articles is simpler in Dutch than in German; whereas the former agrees in number and gender of the following noun, the latter also depends on case.
- b. *Organisational complexity*: the variety of ways of arranging components in a sentence (i.e., rules), for example, constituent order variability (see Table 1.1). The higher the variation in the ways of arranging constituents at the word or sentence level, the higher the system complexity: e.g., Korean—with a fairly flexible constituent order (see 3 in Table 1.1)—is more complex than French or English, which have a fixed SVO constituent order.

Structural complexity

²This thesis focuses on absolute complexity within what Rescher (1998) classified as the “ontological mode of complexity”, and excludes the “formulaic mode”, which would include measures such as Kolmogorov complexity or Minimum Description Length (MDL). Although these complexity metrics are common in the field of language evolution (e.g., Kirby et al. 2015), they are of little use when full grammars are not available/describable, as it is the case during periods of grammatical change in language evolution. For an in-depth account of Kolmogorov complexity see Dahl (2008); Li and Vitányi (1997); Miestamo (2009); and for MDL, see Brighton (2002); Grünwald, Myung, and Pitt (2005); Li and Vitányi (1997).

- c. *Syntagmatic complexity*: the number of rules and/or constituent elements (tokens) in an expression or set of expressions, e.g., morphemes, words or phrases. The more components that appear sequentially, the higher the structural complexity; for example, “Rina gives Mina a present” (5 words) is less complex than “Rina gives a present to Mina” (6 words).
- d. *Hierarchical complexity*: elaborateness of subordination relationships, that is, the number of different levels of constituency (i.e., morphs < words < phrase < sentence < complex sentence). The more levels of the hierarchy of constituency we find in an expression or in a set of expressions, the higher the structural complexity; for example, “The man that Rina met at the party gave a present to Mina”, a complex sentence including a subordinate clause, is more complex than “Rina gave a present to Mina”.

B. Complexity of the form-meaning mapping

- a. *Isomorphism*: directness of the relationship between semantics and morphosyntax. The more computable and motivated by the semantic structure the morphosyntactic structure is, the higher the isomorphism and the lower the complexity of the structure. For example, the phrase “old books and paintings” is ambiguous between two readings: it can be parsed as [[old books] and [paintings]] or as [old [books and paintings]]. The morphosyntactic structure in this case corresponds to the semantic structure; i.e., the meaning the phrase receives corresponds to the hierarchy between the morphosyntactic constituents.
- b. *One-to-one mappings (transparency)*: clarity of relation between meaning and form. Unlike isomorphy, transparency does not concern structure, only reference. The more deviations there are from one-to-one form-meaning mappings in terms of e.g., synonymy, allomorphy, fusion, fission or homonymy, the less transparent the mapping and the higher the system complexity. Generally, we find fewer one-to-one correspondences in fusional languages such as French than in agglutinative languages such as Korean.

1.2.3.1 Measuring absolute complexity

So far in the examples provided I have been quantifying absolute complexity by the number of types within a given feature (system complexity) or the number of component parts within a given expression or set of expressions (structural complexity). This *feature-based* approach (Atkinson 2016) is the most common in cross-linguistic analysis. Feature-based measures are particularly useful for the characterisation of *system complexity* (e.g., definite article paradigms or constituent order variants) as they have the advantage of allowing large cross-linguistic comparisons using widely available typological datasets such as the World Atlas of Language Structures (WALS, Dryer & Haspelmath 2013) or the integrated Typological Database System (TDS, Dimitriadis et al. 2009). *Structural complexity* can also be quantified in a similar fashion from corpora rather than typological databases, i.e., by counting the number of times a specific linguistic feature occurs in an expression or in a set of expressions on average: e.g., hierarchical complexity can be compared between languages by the average number of subordinate relations between the nodes in their derived syntactic trees (e.g. Ferreira 1991).

Although feature-based analyses can be used to make useful (and broad) typological observations (e.g., differences in complexity between languages with high and low percentage of foreign speakers (Bentz & Winter 2013; Good 2015; Lupyan & Dale 2010; McWhorter 2005)), they do not take into account any rules that might condition the choices of one or another type for a given feature. As we saw in the discussion of Table 1.2, Dutch and French both differentiate singular definite articles—in full form—by the gender of the following noun. However, whilst Dutch distinguishes between common and neuter genders (*de* and *het* respectively), French distinguishes between masculine and feminine (*le* and *la* respectively). In common-neuter noun class systems, the common gender is significantly more frequent (approx 75% and 25% in Dutch, see Hulk & Cornips 2006), whilst in feminine-masculine systems, both genders can appear in comparable numbers in the noun lexicon (approx 44% and 56% in French, see Roché 1992). The frequency of each noun class in the lexicon, as well as their recurrence in a natural language sample, will determine the probability of finding each of the singular forms of the definite article variants. In order to take factors such as frequency into account, *at least an extended featured-based* approach is required, whereby complexity could at least be quantified in terms of the ratios between the types and tokens, which takes into account frequency and

richness more broadly. For this type of measure, typological description does not suffice to quantify even system complexity, and linguistic corpora are required. We could then measure the system complexity (paradigmatic, more specifically) of the Dutch and the French definite articles by calculating their type-token ratio, i.e., the number of times each variant type is used divided by the total number of definite articles tokens in the corpus. We can also find measures of this sort for structural complexity in corpus linguistics (e.g. Lu 2010) and production studies in language acquisition (e.g. Gilabert 2007).

Alternatively, *information-theoretic* approaches offer a more comprehensive quantification of complexity than the featured-based approach with the additional advantage of pertaining to a widely established framework—unlike the less articulate extended feature-based approaches discussed. Information theory uses Shannon entropy or entropy-related measures to quantify the amount of uncertainty contained in a system or structure; uncertainty under this account is equated to complexity. Complexity is thus highest in the case of total randomness; the more patterned an object is, the simpler it is (for alternative accounts which exclude randomness as noise and only measure the complexity of patterns, see Gell-Mann 2003; Newmeyer & Preston 2014). Entropy allows the quantification of recurrent patterns or forms taking into account their probability distributions. Going back to our example of the definite article, we can quantify how variable the use of a singular definite article form is in French within a real language sample. For a set of observations S , its entropy is given by

$$H(S) = - \sum_{i=1}^n P(s_i) \log_2 P(s_i), \quad (1.1)$$

where the sum is over the different variants (i.e., *la* or *le*) and $P(s_i)$ is the relative frequency of variant s_i in the set of observations S . If *le* and *la* occur equally often, $H(S) = 1$. If *le* occurs 75% of the time and *la*, 25%, $H(S) = 0.81$. The more variability in the system, the less certain we are that a given form will be used, and the less certainty, the greater complexity. This is an example of what would count as paradigmatic complexity, but entropy can also account for other types of system complexity, and structural complexity: for example, we can quantify how variable word order is (organisational complexity) or the average variability of morphology in individual expressions (syntagmatic complexity).

Entropy can be extended to consider the relationship with other linguistic and extra-linguistic

features. For example, conditional entropy quantifies how variable the use of a singular definite article is in French given information about the gender of the following noun. Given a set of observations S in the set of contexts C , conditional entropy is given by

$$H(S|C) = - \sum_{c \in C, s \in S} P(c, s) \log_2 \frac{P(c)}{P(c, s)}, \quad (1.2)$$

where the sum is over the different variants (i.e., *la* or *le*) and contexts (gender of the following noun, i.e., masculine or feminine); $P(c)$ is the relative frequency of context c in the set of contexts C ; and $P(c, s)$ is the relative frequency of variant s in context c in the set of observations and contexts, S and C . If *le* and *la* always occur in the context of a masculine and a feminine respectively, $H(S|C) = 0$; we can be certain of the definite article that will appear if we know the gender of the following noun. Entropy and conditional entropy capture two different aspects of complexity; the former is a measure of uncertainty of the variant use within definite articles, the latter measures the amount of uncertainty that remains after the gender of the following noun is known. Whilst paradigmatic complexity is higher in French than in English (which only has one definite article), the complexity is predictable by the context and thus the conditional entropy of French and English is the same. Altogether, definite articles are still more complex in French than in English because we require an extra bit of information about the context to predict the variant that will appear. The system of singular definite articles is more complex in German than in English in all respects (see 1.2): not only would entropy be much higher but this would also apply to conditional entropy. The German system of definite articles requires the knowledge of three different linguistic features in order to fully predict the variant use of definite articles in a given context: the gender, the case and the argument role of the following noun.

By knowing the entropy of one variable (e.g., definite articles) and its conditional entropy in a given context (e.g., gender of the following noun), we can also compute the mutual information between these two variables; i.e., the reduction of uncertainty that knowing either variable provides about the other. Given a set of observations S and the set of contexts C ,

mutual information is given by

$$\begin{aligned}
 I(S;C) &= H(S) - H(S|C) \\
 &= H(C) - H(C|S) \\
 &= H(S,C) - H(C|S) - H(S|C)
 \end{aligned}
 \tag{1.3}$$

where $H(S,C)$ is the joint entropy of S and C . In the case of singular definite articles in French, *le* always appears with masculine nouns and *la* with feminine nouns, therefore $I(S;X) = 1 - 0 = 1$. Additionally, mutual information can be used to quantify the strength of association between meanings and forms, and thus measure the transparency of form-meaning mappings. Mutual information is a very robust measure of association but it does not provide information about the direction of the association, either positive or negative. This means we cannot know whether a given meaning is the referent of a given form, or alternatively the meaning never appears when a given form is used. Moreover, the maximum mutual information is the joint entropy of the two variables and therefore mutual information scores are not comparable between different types of systems. In order to solve this comparability problem, we can use the mutual information related measure known as Jensen-Shannon distance (Endres & Schindelin 2003), a distance metric between two probability distributions bounded by 1 (i.e., maximum distance is 1; minimum, 0). Alternatively or complementarily, we can deviate from information-theoretic measures and use correlation coefficients to extract the strength as well as the direction of the relationship between the occurrences of a given meaning and those of a form. Although correlation coefficients are not as robust a measure of association in comparison to mutual information and they can only detect linear relationships, they provide information about the direction of the relationship as well as clear minimum and maximum strengths of association. Correlation coefficients range between 1 and -1 , where 1 is total positive linear correlation, 0 is no linear correlation, and -1 is total negative linear correlation. The stronger the positive association, the more transparent the form-meaning mapping; a coefficient of 1 would suggest a strictly one-to-one form-meaning mapping. We can also use correlation coefficients to quantify isomorphism (i.e., systematic mapping between morphosyntactic structure and semantics) in a language sample, for example, by computing the correlation between a matrix of morphosyntactic-distances and another of semantic-distances of the expressions comprised in

the sample (Kirby et al. 2008, 2015). The higher the correlation between distances across morphosyntactic structures and semantic structures, the more isomorphic the linguistic structure is.

Information-theoretic measures can also account for descriptive complexity, also known as algorithmic entropy or Kolmogorov complexity, which I have excluded from the taxonomy for the sake of simplicity as it will not be used in this thesis (for an in depth account of Kolmogorov complexity, see Dahl 2008; Li & Vitányi 1997; Miestamo 2009). Kolmogorov complexity is calculated by the minimum number of bits from which a given system can be reproduced, or in other words, the length of the shortest possible description—in some sort of universal description language. Unlike entropy-related measures, it is not contingent on any probability distribution that generates the data, thus it does not take into account frequency effects either. Kolmogorov complexity is a common metric of complexity, however it is very dependent on the description language used and, more problematically, it is uncomputable (Li & Vitányi 1997). It can nevertheless be approximated using the Minimum Description Length principle (MDL), which states that the best description for a given set of data is the one that leads to the best compression of the data as well as the description of the data (for an in-depth account of MDL, see Brighton 2002; Grünwald et al. 2005; Li & Vitányi 1997). This measure of descriptive complexity is common in the study of grammar complexity in computational models of language evolution (e.g. Brighton et al. 2005; Kirby et al. 2015). However, as Deutscher (2009) points out, descriptive complexity is not suitable for the evaluation of linguistic systems whose grammar cannot be fully described, which is the norm in the messy and limited behavioural product from experimental models.

1.2.4 Relationship between absolute and relative complexity

Absolute and relative complexity can, and often do, go hand in hand. If we compare the most explicit criteria that have been previously proposed for cross-linguistic comparison from each approach—relative and absolute—we realise that they are in many respects similar (Miestamo 2008).

Kusters (2003, 2008) establishes three criteria whereby relative complexity for the “generalised user” is measured: *economy*, *transparency* and *isomorphy*. Economy is proportional to the number of categories that can be overtly encoded; the fewer categories a language allows,

the more economical it is. Economy here is proportional to system complexity, in particular, to paradigmatic complexity. Transparency refers to the clarity of form-meaning mappings; the more one-to-one form-meaning mappings, the more transparency. Lastly, isomorphy is measured by the directness of the relationship between syntax and semantics; the more computable the semantics are from the syntax, the more isomorphic the linguistic structure is. The author shows that violations of these criteria cause difficulty for L2 learners and thus are interpreted as complexity. These criteria for relative complexity perfectly align with those in the taxonomy of absolute complexity in section 1.2.3. And although the study focuses on inflectional morphology, these criteria can be easily generalised to other aspects of a grammar.

McWhorter (2001) argued that pidgin/creole grammars are simpler than those of older languages in terms of absolute complexity according to three criteria he developed further (McWhorter 2007): *overspecification*, *structural elaboration* and *irregularity*. Overspecification refers to the number of grammatical distinctions overtly and obligatorily encoded; the more distinctions a language makes within a given domain (e.g., definite articles) the higher its paradigmatic complexity. Structural elaboration involves “the number of rules (in phonology and syntax) or foundational elements (in terms of phonemic inventory) required to generate surface forms” (McWhorter 2007 p. 29); structural elaboration would be proportional to the descriptive complexity of a grammar, and in our computable taxonomy would mainly comprise both organisational and syntagmatic complexity. Finally, the criterion of irregularity concerns the amount of irregularity and suppletion; the longer the list of separate items in a system, the higher its complexity. The overspecification and irregularity criteria in McWhorter (2007) are close to Kusters’ (2003; 2008) economy and transparency. The more overspecification, the less economy; and the more irregularity, the fewer one-to-one form-meaning mappings. The fact that their criteria are so similar might be due the overlap of the object of study; i.e., second language adult learning (Kusters 2003, 2008) and pidgins/creoles (McWhorter 2007), which are languages that have evolved from populations with a majority of adult learners. Following these criteria, Miestamo (2006a) proposed two very general principles to compute absolute complexity, which the author claimed would apply to older as well as newer languages: the principle of *Fewer Distinctions* and the principle of *One-Meaning-One-Form*. In the same way, the fewer distinctions and the more transparency, the simpler the language.

In sum, the criteria to assess relative and absolute complexity in cross-linguistic studies

overlap noticeably. In the following sections I will discuss models of language evolution which suggest a causal relationship between absolute linguistic and competing pressures in language learning and use to minimise relative complexity. These models provide suggestive evidence for the claim that when languages are shaped by the pressures imposed on the user in learning and communication, under the taxonomies discussed, linguistic systems tend to become simpler in absolute terms. However, further work is required to comprehensively explore such claims across different aspects of absolute complexity in morphological and syntactic systems.

1.3 Linguistic complexity as an evolving variable

It is common practice in cross-linguistic analyses to assess the complexity of languages as static systems—as we have done so far in our examples of absolute complexity. However, languages are dynamic systems product of cultural evolutionary processes, and as such they are constantly adapting to pressures imposed on the learner/user during transmission. We might study the dynamics of linguistic complexity over various time-scales: we can explore the emergence of complexity in language formation and its modification in language change, or its development during language learning and its alteration during language use in communication.

Across time-scales, both complexification and simplification processes can be explored. In language emergence the focus is usually on the shift from simpler to more complex structure (Dahl 2004; Galantucci & Garrod 2011; Kirby et al. 2008). However, complex structure can develop from simpler systems or from unstructured and/or chaotic systems. In the latter case, the emergence of patterns in structure is accompanied by a reduction of system complexity—i.e., randomness decreases. In regard to historical language change, processes such as grammaticalisation or language contact can also be studied with a focus on whether they increase or decrease the complexity of a system (Heine & Kuteva 2007; Trudgill 2011). For instance, contact between languages has been documented to result in “additive” borrowing—in which new features are acquired in addition to existing features instead of replacing them (Kuteva 2008; J. Nichols 1992), but they also result in the reduction of redundancy and in the increase of regularity and transparency (Sasse 2001; Trudgill 2011)—especially in extreme contact situation which lead to the formation of pidgin/creole languages (McWhorter 2007; Mühlhäusler 1977). During language acquisition, learners start out with simpler systems—manifested in

e.g., simpler morphosyntax and poorer lexicons—which increase in complexity as they match the target languages more closely; but are the systems simpler only because they have not been fully acquired or do we find additional processes of simplification? There is a large body of work that provides evidence for simplification behaviour (i.e., regularisation in particular) in early stages of L1 acquisition (Fraser, Bellugi, & Brown 1963; Newport 1999; Ross & Newport 1996; Singleton & Newport 2004) and in L2 interlanguages (Ellis 2013; Kusters 2003; Richards 1974, 1975; Touchie 1986).

The study of linguistic complexity as a dynamic variable during communicative interaction has not often been addressed directly; whilst many researchers subscribe to the idea that languages adapt to communicative needs (see contributions in Ellis & Larsen-Freeman 2009; Massip-Bonet & Bastardas-Boada 2012), few studies explore how linguistic complexity is modified accordingly during the course of communicative interaction (e.g., Atkinson 2016; H. H. Clark & Wilkes-Gibbs 1986; Fehér, Wonnacott, & Smith 2016; Fussell & Krauss 1989). The few available suggest that although speakers initially tend to produce longer and more redundant expressions to ensure their comprehension (Fussell & Krauss 1989), as linguistic conventions develop during interaction, we observe a drive towards shorter utterances and more regular systems (H. H. Clark & Wilkes-Gibbs 1986; Fehér, Wonnacott, & Smith 2016).

The studies in this thesis will explore how relative complexity (mainly user-general) affects absolute complexity across these different time scales; i.e., how language learning and use impact the complexity of linguistic systems, during a systems's development and evolution. Understanding any causal relationship between relative and absolute complexity across time-scales requires us to address what Kirby (1999) referred to as *the problem of linkage*, i.e., how individual behaviours in language learning and use shape the complexity of the way a language is organised and structured at the population level. Cultural transmission provides a solution to this problem (Kirby 1999). Languages are culturally transmitted through a repeated cycle of learning and communicative interaction, also known as *iterated learning*; therefore it makes sense to posit that languages adapt over cultural time to maximise learnability as well as communicative effectiveness, altogether maximising their efficiency (Beckner et al. 2009; Brighton et al. 2005; Kirby 1997; Regier et al. 2015; K. Smith & Kirby 2012; Steels 2012). Through cultural transmission, the product of the learners/users' biases during language learning and to communication accumulates over repeated usage and over cultural time, ultimately

shaping a system of behaviour shared at the population level (Brighton et al. 2005; Kirby, Griffiths, & Smith 2014). How learnable features are and how effective they allow communication to be will determine their prevalence in the linguistic system, eventually leading competing features which are less fit to remain a minority or dissipate.

Computational and experimental iterated learning models and further signalling games provide extensive support for these claims (e.g. Baronchelli, Loreto, & Steels 2008; Brighton et al. 2005; Christiansen 2000; Kirby & Hurford 2002; Kirby et al. 2015; Nowak & Baggio 2016; Regier et al. 2015; Spike, Stadler, Kirby, & Smith 2017; Steels 1995; Verhoef 2012). These studies often focus on how linguistic structure evolves over generations of learners/users—from scratch or from unstructured systems—as a trade-off between cultural pressures at play in language transmission (e.g., Brighton et al. 2005; Kirby & Hurford 2002; K. Smith 2004; Steels 2012; Theisen-White, Kirby, & Oberlander 2011). Only recently are we observing an increasing interest in answering how *complexity* in linguistic structure evolves (e.g., Atkinson 2016; Mufwene et al. 2017; K. Smith et al. 2017; Tinitis, Nölle, & Hartmann 2017; van Trijp 2016; Winters 2017). The current thesis aims to contribute to this latter question. Having discussed this widely accepted solution to the problem of linkage in the field of evolutionary linguistics (i.e., cultural transmission), we now focus on defining the pressures involved in communication and learning in more detail, and what kind of short-term behaviour they trigger to ultimately understand how these might impact on the complexity of the morphology and syntax of a language.

1.3.1 Communicative effectiveness and learnability in language transmission

1.3.1.1 Communicative effectiveness

Effective communication is first and foremost dependent on a shared system of conventions between users (Lewis 1968; Schelling 1960). Every communicative interaction requires speakers to convey a meaning with a signal and hearers to arrive at an interpretation. In order to meet this requirement users need to coordinate to align on a shared system of conventions, established through the repeated use of the same expression to refer to the same thing; such routinisation at the same time facilitates the identification of new as opposed to old interpretations for any given expression (E. V. Clark 1988). Additionally, effective communication

requires that messages are expressive enough to unambiguously convey a meaning in a given context; an ambiguous message would hinder the task of the hearer to arrive at its intended meaning. There are thus two pressures at play for effectiveness in communication, a pressure for coordination—to facilitate linguistic alignment between interlocutors—and a pressure for expressivity—to ease comprehension of the meaning to be conveyed.

1.3.1.2 Learnability

Learnability is often described in terms of robust transmission through a bottleneck (Brighton et al. 2005; Kirby 2002; K. Smith, Brighton, & Kirby 2003). Languages are productive and productively interpretable systems; however they must be learned from messy and relatively limited input, which posits a challenge to the learner. This mismatch between the magnitude of the system to be transmitted and its medium of transmission is referred to as a *learning bottleneck* (Kirby 2002; K. Smith & Kirby 2008; Spike et al. 2017). The learner needs to learn a grammar (L1 or L2) that generates a potentially infinite amount of data regardless of this bottleneck in transmission (Chomsky 1980). How can languages then be robustly transmitted? We find two complementary but often encountered answers to this question in the literature: generative accounts propose that humans are biologically endowed with a *language faculty* that complements the “impoverished input” (Chomsky 1986); and cultural evolutionary accounts claim that language adapts over time to maximise its learnability making its instances as generalisable as possible (Brighton et al. 2005; Zuidema 2003). Under the latter account, languages are generalisable enough to allow productivity with the acquisition of a finite lexicon and a finite set of rules (i.e., grammar); crucially, they evolved this way due to the accumulated behaviour product of a bias for learnability in language acquisition. Learnability is generalisability: the *more generalisable* the lexicons and grammars, the more learnable (i.e., easy to induce) they will be from the limited amount of data the learner is exposed to. These claims are supported by experimental and computational iterated learning studies (e.g. Brighton et al. 2005; Christiansen & Chater 2008; Kirby et al. 2008, 2015; Regier et al. 2015; Wonnacott & Newport 2005; Zuidema 2003).

A bias for generalisability has also been shown to be at play even in the absence of a learning bottleneck from limited data; i.e., even when all the data is available, learners tend to regularise any inconsistencies they might find, making the system more generalisable (Cul-

bertson, Smolensky, & Legendre 2012; Hudson Kam & Newport 2009; K. Smith & Wonnacott 2010). Regularisation behaviour during individual learning and use (without communication) has been proposed to be due to a bottleneck in memory instead, which could affect the encoding of variants and their relative frequencies in training, and/or in variant retrieval during production. Altogether, these studies suggest a bias for learnability which leads to the reduction of a system's complexity (Brighton et al. 2005; Kirby et al. 2008, 2015); I refer to it as a bias for learnability (Kirby et al. 2015) or generalisability (Winters 2017), but it is also referred to as *compressibility* or *simplicity* bias. The reduction of complexity is quantified in terms of algorithmic complexity (i.e., compressibility) in computational iterated learning models—which allow for a full description of a fabricated grammar (Brighton et al. 2005; Kirby et al. 2015), and in terms of paradigmatic and organisational complexity in laboratory experiments (Atkinson 2016; Culbertson et al. 2012; Kirby et al. 2008; K. Smith & Wonnacott 2010; Winters, Kirby, & Smith 2015). Generally, we can conclude that the higher the generalisability of a system, the lower its uncertainty and thus the lower its *system* complexity. The same conclusion cannot be drawn about structural complexity, which is not usually the focus of study and can be orthogonal to system complexity (Kirby et al. 2015). For example, the system complexity of the English verbal system would lower if there were no irregular forms as it would contain less variability within morphological systems. However, the substitution of irregular for regular forms could lead to an overall increase in structural complexity; e.g. *runed* and *swimed* are composed of two concatenated morphs whilst *ran* and *swam* only contain one (suppletive) morph.

A pressure for generalisability is also present in coordination during communication, where learning also takes place. Messages need to be produced and interpreted productively, thus the conventions interlocutors align to during interaction need to be generalisable. Conventions not only need to solve the immediate task at hand to achieve communicative success but they also need to be able to be reused to solve future problems—which constitutes a bottleneck in communication. If conventions are not generalisable to new data and to new situations (e.g., new context, new interlocutors), the probability of making successful predictions, and thus communicating successfully, decreases (Winters 2017). Laboratory experiments exploring specifically the role of communication in the emergence of linguistic structure and regularisation behaviour have shown that efficient and generalisable linguistic systems are indeed developed during in-

teraction (Fehér, Wonnacott, & Smith 2016; Winters, Kirby, & Smith 2018).

1.3.1.3 Efficient linguistic systems: the interplay between learnability, expressivity and coordination

An *efficient* (not only effective) linguistic system requires the coordination of linguistic conventions which are both expressive and learnable (equated to generalisable hereafter) (Winters 2017). A language only shaped by a generalisability pressure can result in degenerate systems whereby all possible meanings are encoded by the same expression. Such a system would be maximally learnable but also maximally ambiguous, which would not allow its users to discriminate between meanings—given that disambiguating all meanings extra-linguistically is unattainable (Levinson 2000). On the other extreme, a language only shaped by an expressivity pressure would lead to holistic languages, whereby each meaning to be conveyed is expressed by a different expression which cannot be further divided into meaningful units. A holistic system is maximally expressive and allows its users to communicate successfully; nevertheless it would not allow learners to use this linguistic knowledge and generalise it to novel meanings, which will set a cap on productivity—unless learners are exposed to the full language and their memory resources allow them to acquire it. Whereas the absolute complexity of the form would be minimal in a degenerate language and higher in a holistic language, the complexity of form-meaning mappings would be maximal in degenerate systems and minimal in holistic.

One solution to the interplay between learnability and expressivity pressures is to generate a language with isomorphic and transparent form-meaning mappings, which will maximise interpretability and maintain system complexity as low as possible (i.e., the number of components to be learned), altogether minimising the effort of unambiguously conveying and interpreting a message. Supporting evidence comes from computational and experimental studies modelling iterated learning which have shown that compositional structure—whereby the meaning of the form is derived from the meaning of its constituent parts and the way they are combined (Szabó 2012)—evolves from the trade-off between learnability and expressivity pressures (e.g. Brighton 2002; Kirby & Hurford 2002; Kirby et al. 2015; Regier et al. 2015). Compositional structure, along with an unbounded hierarchical structure (Chomsky 1965), allows for productive (generalisable) and productively interpretable (expressive) languages, and it only requires its users to learn a finite lexicon and a finite set of combinatorial rules (i.e.,

a grammar) (K. Smith & Kirby 2012). Therefore whereas the emergence of compositional structure might be accompanied by an increase in structural complexity, it is in the interest of system complexity and productivity; it licenses users to “make infinite use of finite means” and potentially communicate about anything with each other. So far only experimental evidence is available for the evolution of compositional structure without hierarchical structure, but hierarchy is theorised to emerge along the expansion of the worlds to communicate about (e.g., see Kirby 2002; Mufwene 2012), to precisely license more productive and productively interpretable languages.

Another source of efficiency in linguistic systems comes from the trade-off between expressivity and effort expenditure in terms of time and word/utterance length (respectively, see the principles of Economy, Minimise Forms and Least Effort in Haiman 1983; Hawkins 2004; Zipf 1949). According to Zipf, this principle (amongst other things) can explain difference between word length as a function of their frequency: by reducing the length of the most frequent words, speakers spend on average less effort producing their utterances (for behavioural evidence of Least Effort, see Kanwal, Smith, Culbertson, & Kirby 2017). This principle has been reformulated also by H. H. Clark and Wilkes-Gibbs (1986) into the principle of *least collaborative effort* to include the collaborative aspect in the establishment of linguistic conventions during communicative interaction. This principle states that interlocutors exploit the context to minimise the *total* effort spent during interaction, i.e., in both production and comprehension (H. H. Clark & Schaefer 1989; H. H. Clark & Wilkes-Gibbs 1986; Davies 2006). Support for least collaborative effort comes from experimental studies showing that speakers tend to reduce structural complexity over repeated interactions—as a result of the elimination of redundant information to the hearer and/or the context (H. H. Clark & Wilkes-Gibbs 1986; Fehér, Wonnacott, & Smith 2016; Hupet & Chantraine 1992; Kanwal et al. 2017; Krauss & Weinheimer 1964, 1966; Winters et al. 2015, 2018). At the same time, most of these studies further connect *least effort* to learnability: speakers produce overall more transparent and more regular systems during communicative interaction than they would in isolated production, without the need to communicate (Brennan & Clark 1996; Fehér, Wonnacott, & Smith 2016; Fox Tree 1999; Fussell & Krauss 1989; Garrod & Anderson 1987). At the end of the day, both learnability and utterance shortening help users to minimise the effort of unambiguously conveying a message.

1.3.2 Functionality in the evolution of linguistic complexity

So far I have argued that linguistic complexity can be influenced by the pressures imposed on the user during language transmission, both in learning and communication. The cultural product stemming from these pressures adapts over repeated usage, and potentially over cultural time, to maximise the language's learnability and communicative efficiency. Moreover, I have discussed how systems and form-meaning mappings can become simpler in the process. As a matter of fact, simplification is a common process in natural language change and evolution more broadly. Over time, irregularities are regularised, transparency increases and grammatical distinctions are lost; e.g., in the evolution of English, irregular verbs and plurals such as *holp* and *kine* have been replaced by *helped* and *cows* and case morphology was lost to a more rigid word order to indicate semantic roles. However, concluding that languages are shaped by the pressures at play in language transmission—during language learning and use—does not entail that these will always lead to language change in the long run or that they are the only trigger of language change. All the evidence discussed suggests is that language change *can* be a product of these selective pressures in language transmission, and thus that language learning and use can partly explain the evolution of languages and their complexity. Given the cross-linguistic variation of complexity in similar features and domains (e.g., the number of distinct definite articles or word order flexibility), it is hard to make the case for strict functionality (Dahl 2004; Gil 2009).

As pointed out by Gil (2009), languages are generally more complex than required to assure learning and communication. The author argues that a language which is purely isolating, which has no distinct syntactic categories, and whose form-meaning mappings are purely associational would suffice; he calls this an *isolating-monocategorical-associational* (IMA) language. This language would not require any type of linguistic structure (including word internal structure), would only count with an undefined single syntactic category, and it would have no semantic interpretations dependent on linguistic structure. And since IMA languages do not exist, Gil (2000) concludes that language structure is not functional but simply the product of system-internal self-organisation processes. This conclusion assumes that any adaptation to the user's needs leads to simplification ubiquitously; and although as previously discussed there exists an overarching bias towards simplicity, we cannot conclude that it applies to all aspects

of linguistic complexity or that it is the only outcome from the interaction between different pressures in cultural transmission and across contexts. For example, we have previously mentioned that in the emergence of productive and productively interpretable languages, structural complexity necessarily increased in the interest of the system's economy and productivity—i.e., to maintain learnable finite sources whilst allowing users to “make infinite use” of them (see also [Mufwene 2013](#)). Such productivity would be compromised in an IMA language, both in production and interpretation; an IMA language would be heavily underspecified and too dependent on contextual information, therefore it would put a cap in productive interpretability. Different linguistic contexts can indeed lead to differences in communicative contexts and pressures ([Trudgill 2011](#); [Wray & Grace 2007](#)). We can imagine that an IMA language could be effective in small populations where interlocutors shared a long history of communicative interaction that can provide them with the common ground required to overcome underspecification. However, as soon as the communicative network and the meaning space complexified, linguistic structure would develop as it has around the languages of the world (for an experimental proof of concept, see [Winters et al. 2018](#)).

Languages evolve cumulatively through cultural transmission, with different components being added, reduced or modified at different times and contributing to different domains. The cumulative fashion in which languages are transmitted between users and from generation to generation highlights not only the importance of the pressures imposed on the general user but also the importance of historical contingencies. The evolutionary paths of linguistic complexity are constrained by historical events (and their linguistic and communicative contexts) that are often random. Thus the emergence of complexity is guaranteed from sources other than adaptation as well. Under this view, and not taking languages as static systems, [Gil's \(2009\)](#) denial of any adaptation to the learners'/users' biases is unfounded. Not all features in languages respond to learnability and communicative effectiveness, but languages do not complexify and simplify by drift alone either. Language learning and use play an important role in the cultural transmission and selection of linguistic features, and it is worth further exploring how individual biases impact what aspects of the complexity of linguistic systems over repeated learning or communicative episodes, and over cultural time.

1.4 Research focus and methodology

In their introduction, Mufwene et al. (2017) write that “students of linguistic complexity must explain the consequences of thinking of languages as complex adaptive systems” (p. 14). This thesis aims to contribute to such an explanation. Thinking of languages as complex adaptive systems requires us to focus on the processes and mechanisms involved in their transmission as cultural products; it compels us to explore how absolute complexity is determined by user-related complexity. In the previous section I have discussed cognitive biases and cultural pressures which have been proposed to be at play in language transmission. Whereas there is plenty of evidence on how the interaction of these pressures and biases shape linguistic structure during language learning and use, and over cultural time, there is little evidence of how they shape the complexity of such structure over different time scales and across different linguistic domains.

My contribution to the field is thus to directly test the conditions under which complexity in morphology and syntax emerges, spreads, and reduces during language learning and use, and over generations of learners/users. I do so using laboratory experiments with young adults. Laboratory settings allow us to observe trajectories of change, and test *causal* hypotheses while controlling variables that are difficult or impossible to control in the real world (Cangelosi & Parisi 2012). Thus unlike corpus and cross-linguistic analysis, laboratory experiments let me test predictions about causal relationships in a controlled environment where I can operationalise complexity, the learning and communicative contexts and the different time scales. This advantage also comes with its limitations: the data obtained from experiments will not reflect as closely the complexity of real-world language, its environment, or the dynamics of language change and formation, and it will lack the power and the richness of large corpus and cross-linguistic analysis (see section 1.4.1 for further discussion).

For the studies in this thesis I will employ techniques from iterated learning (Esper 1966; Kirby et al. 2008, 2015), interaction studies (within experimental semiotics and pragmatics, e.g., H. H. Clark & Wilkes-Gibbs 1986; Galantucci 2005; Garrod & Anderson 1987; Krauss & Weinheimer 1964), and statistical learning (Culbertson et al. 2012; Hudson Kam & Newport 2005; G. A. Miller 1958). All these paradigms (except for interaction studies in experimental pragmatics) use artificial languages to provide direct behavioural evidence of individual biases

in learning and use; artificial languages (novel meanings and forms) facilitate the isolation of specific aspects of natural languages and observe under what conditions these change. I will use artificial languages in the experiments encompassed in Chapter 2 and 4, however, in Chapter 3, I address the effect of communicative interaction without any effect of language learning and the only way to do that is by exploring natural language production. Further description and motivation for the specific paradigms will be provided for each study in Chapters 2–4.

1.4.1 Methodological commitments

Previous work has pointed out the limitations of artificial language learning and iterated learning as general models of language acquisition and change (Beckner & Wedel 2009; Croft 2004; Niyogi & Berwick 2009; K. Smith 2009; K. Smith et al. 2017). K. Smith (2009) and K. Smith et al. (2017) note that while iterated learning provides a powerful tool for exploring how biases in language learning and use shape linguistic structure, in real populations those biases are fed in to a population dynamic—with complex social networks—whose consequences are largely not understood (see also Kerswill & Williams 2000; Lupyan & Dale 2015; Niyogi & Berwick 2009). Further, the authors show that the relationship between individual biases of learners/users and universals in linguistic structure is not transparent: not only can strong effects in languages be due to very weak individual biases but even very strong biases can be completely invisible at the level of languages (K. Smith et al. 2017). In real populations, learners learn from multiple models and interact with multiple speakers, and the outcome of cultural transmission is not necessarily simply determined by individual biases—whose strengths might also differ between individuals within a population (Navarro, Perfors, Kary, Brown, & Donkin 2017)—but also by transmission factors such as the quantity and the quality of data learners and users receive and interact with (Beckner & Wedel 2009; Niyogi & Berwick 2009; K. Smith et al. 2017). Altogether, previous work does indeed suggest that caution must be taken when extrapolating from language learning and cultural evolution in convenient miniature language learning tasks and one-individual/pair iterated learning chains to larger populations.

I make no strong claims about the nature of the individuals (learners and/or users) driving natural language change in this thesis either; however, the studies contained ascribe to a model of language change and formation in which both acquisition and usage are crucial drivers of change. There is no reason to assume that language change is driven exclusively via acquisition,

or exclusively via usage (Beckner & Wedel 2009; Kirby et al. 2015; Winters et al. 2018).

Language acquisition is undoubtedly an important factor in understanding natural language change: the language system that a learner acquires might differ from that of their input models (e.g. parents). This has been claimed to be a significant part of language change (e.g., Crain, Goro, & Thornton 2006; Lightfoot 2010). Moreover, studies on naturally emerging sign languages and creole/pidgin formation provide further evidence for the role of acquisition in language change (Hudson Kam & Newport 2005; Newport 1999; Senghas & Coppola 2001; Senghas, Kita, & Özyürek 2004). On the other hand, usage-based approaches leave open the possibility that imperfect learning during (at least L1) acquisition might not be the primary mechanism of some types of language change. Although language learning abilities deteriorate with age, mature grammars are not immutable: language usage as a mechanism of change has been proposed to involve continuous, gradual adjustments to language structure across the lifespan of each individual (Bybee & Slobin 1982; Kerswill 1996; Sankoff & Laberge 1974). There is evidence suggesting that adult speakers adopt ongoing changes in their language (e.g., Harrington 2006), that adults have innovated systematic grammatical conventions in L2 speech communities (Bentz & Winter 2013; Lupyán & Dale 2010; Sankoff & Laberge 1974), and that adults most likely originate grammatical features that are acquired late by children (Bybee 2009). Moreover, there is also evidence that many errors produced by young children do not have a direct reflection in the current direction of change in English (e.g., Bybee & Slobin 1982); it has been proposed that this is the case because children do not constitute an influential group for other learners and users (Kerswill 1996).

Further computational and experimental studies support this dual-mechanism approach to language change in which both acquisition and usage are important drivers. In Beckner and Wedel (2009), the authors use iterated learning computer models to tease apart these two mechanisms of language change; they conclude that—in their case study of regularisation of irregular morphology—both usage and acquisition can be theoretically possible mechanisms of change. Although iterated learning started by having language acquisition at the core of a model of language change in the evolution of structure, the latest work has incorporated both inter- and intra-generational change, with special attention being paid to the the role of usage in interaction with learning (Kirby et al. 2015; Silvey, Kirby, & Smith 2015; Theisen-White et al. 2011; Winters et al. 2015, 2018). As mentioned above, Kirby et al. (2015) provides evidence for

the claim that acquisition and use might sometimes influence language in a different direction, and that their influences may often compete (Kirby et al. 2015). Given the evidence reviewed, a full account of language change will thus acknowledge that the interactions between usage and acquisition are quite complex; under different circumstances the contributions of each of these mechanisms may differ, and in many occasions they may amplify or compete with one another.

1.5 Outline of the thesis

This thesis comprises three experimental studies supporting the hypothesis that the cultural and cognitive pressures acting on language users during learning and communication—for learnability, expressivity and coordination—determine (at least partially) the emergence, maintenance and reduction of absolute linguistic complexity depending on whether or not it maximises learnability and communicative effectiveness, altogether increasing the system’s efficiency.

In Chapter 2, I report the first set of iterated learning experiments to investigate the evolution of complexity in compositional structure at the word and sentence level from holistic, unstructured systems. In Experiment 1 (section 2.2) I demonstrate that a complex meaning space paired with a learning bottleneck in transmission and a pressure for expressivity without explicit communication can result in the emergence of compositional hierarchical constituent structure. This structure includes fixed combinatorial rules of word formation and word order. Compositional hierarchical structure grants a productive and productively interpretable language and only requires learners to acquire a finite lexicon and a finite set of combinatorial rules (i.e., a grammar). The shift from unstructured systems to such linguistic structure entails (as predicted) an increase in structural complexity (both syntagmatic and hierarchical), in the interest of productivity, low system complexity and transparent and isomorphic form-meaning mappings. In Experiment 2 (section 2.3), I show that by combining a learning bottleneck in transmission with communicative interaction facilitates the evolution of compositional hierarchical structure. This supports the claim that communicative interaction cannot be reduced to a pressure for expressivity alone, and that coordination plays an important role in the establishment of linguistic conventions. A third experiment (section 2.4) further corroborates the results

of the previous experiments by showing that if coordination between interlocutors is hindered, so is the evolution of linguistic structure.

Chapter 3 contains Experiment 4 (section 3.5), where I address the unique effect of communicative interaction on absolute linguistic complexity by removing language learning completely. Speakers use their native language to express novel meanings either in isolation or during communicative interaction. I demonstrate that, even in this case, communicative interaction leads to more efficient and simpler linguistic systems: interlocutors produce more productive and transparent morphological lexicons (i.e., lower paradigmatic complexity).

These first two chapters provide support for the claim that morphological and syntactic complexity is shaped by an overarching drive towards simplicity to maximise learnability as well as communicative efficiency. Chapter 4 reports a series of experiments (Experiments 5–7, sections 4.3, 4.4 and 4.6) assessing the uniformity of this simplicity bias across different linguistic levels. Studies in natural language learning and in pidgin/creole formation suggest that while morphological variation seems to be highly susceptible to regularisation, variation in other syntactic features, like word order, appears more likely to be reproduced. I test this experimentally by comparing regularisation of unconditioned variation across morphology and word order in the context of artificial language learning. Variation in morphology and word order are quantitatively comparable but represent two aspects of system complexity, i.e., paradigmatic and organisational complexities respectively. I show that language users in fact regularise unconditioned variation in a similar way across linguistic levels, suggesting that the simplicity bias may be driven by a single, non-level-specific mechanism.

A final chapter provides a summary of the work presented and discusses its contributions to the study of language evolution.

Chapter 2

The cultural evolution of complex compositional structure in the laboratory

2.1 Introduction

Human languages possess a highly productive compositional structure exclusively attested in our species (Collier, Bickel, van Schaik, Manser, & Townsend 2014; Engesser, Ridley, & Townsend 2016; Lachlan & Nowicki 2015; Ouattara, Lemasson, & Zuberbühler 2009; Suzuki, Wheatcroft, & Griesser 2016; Yip 2006). Such a linguistic structure is supported by a remarkable combinatorial capacity for generating an unbounded number of different linguistic signals to communicate complex meanings (Chomsky 1965; Hockett 1960b). These linguistic signals are constructed by recombining reusable meaningless units (phones) to form meaningful units (morphs) which further recombine recursively to form complex meaningful units (from words to sentences to discourse) (Hockett 1960b; Martinet 1967). At the same time, the meaning of such complex units is derivable in a predictable way from the meaning of their subunits and a language's grammar, a property of languages known as compositionality (Pagin & Westerståhl 2010; Szabó 2012). And although grammars differ between languages, the existence of such productive power provided by compositional structure is universal to all languages.

How did such characteristic structure evolve? Evolutionary linguists have effectively stud-

ied it as a product of cultural evolution (Brighton et al. 2005; Christiansen & Chater 2008). Languages are culturally transmitted through a repeated cycle of learning and communicative interaction. These two aspects of cultural transmission impose two interacting pressures that potentially shape the evolution of linguistic structure: a pressure for languages to be learnable (for ease of acquisition) and a pressure for languages to be expressive (for communicative effectiveness). Compositional structure allows language to be expressive and learnable, i.e., language users can communicate potentially about anything “making infinite use of finite means” (i.e., a lexicon and a grammar) (Chomsky 1965).

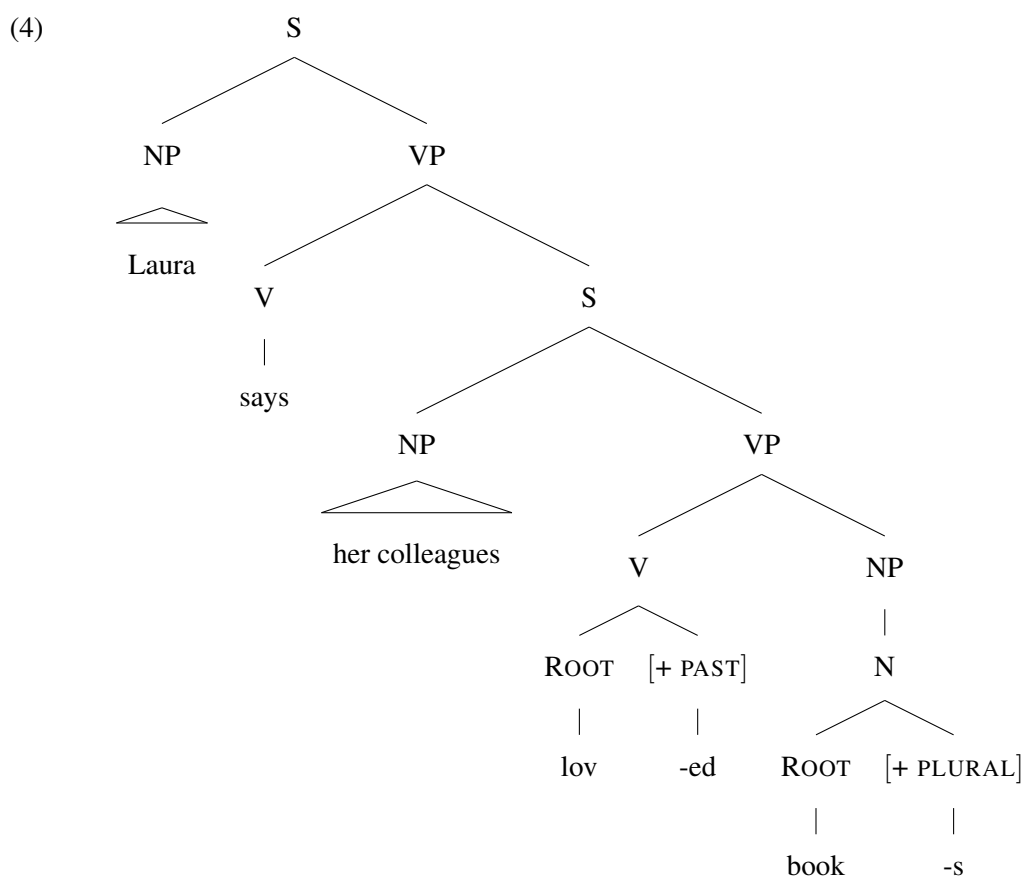
Several experimental models of iterated learning have shown that basic compositional structure emerges from the trade-off between learnability and expressivity pressures in cultural evolution (Kirby et al. 2008, 2015); in these experiments, languages evolve in which simple forms map to simple meanings and these combine by means of concatenation, without further syntactic or semantic structure. In this paper I show how the same mechanisms involved in the evolution of such basic compositionality can lead to richer morphosyntactic structure by increasing the complexity of the meanings to be conveyed. More specifically, in this study I show that hierarchical constituent structure and argument structure (marked via word order rules) can also result from the need for languages to be learnable as well as expressive.

2.1.1 Structure in natural languages

Human languages are productive and productively interpretable communicative systems. Such productivity is facilitated by the interaction of (at least) two properties that linguistic structure often exhibits: compositionality and hierarchical constituent structure.

Compositionality is a property of a language’s semantics relative to its syntax by which *the meaning of a linguistic expression is derivable in a predictable way from the meaning of its constituent parts and the way they are combined* (Pagin 2012; Szabó 2012). In other words, compositionality assures the prediction of a meaning given the form. Any finite system would be trivially compositional because a list of form-meaning mappings can make the predictions, but language is a non-finite system and non-finite lists are impossible. Instead, in natural languages form and meaning are systematically isomorphic, i.e., they have a structure-preserving one-to-one correspondence. Isomorphism thus requires structure in form (i.e., syntax) as well as in meaning (i.e., semantics), and this in human languages is hierarchical constituent structure.

Through syntactic recursion, sentences in natural language can be organised into a hierarchy of constituents—known as constituent structure (Chomsky 1957), where each constituent higher in the hierarchy is built by recursively combining constituent units from lower levels: sentences are built from other sentences and/or phrases; phrases, from other phrases and/or words; and words from other words and/or morphs. The derivation of the sentence *Laura says her colleagues loved books* provided in (4) illustrates these different levels of constituency. Higher levels of constituency (i.e., sentences and phrases) demonstrate greater productivity, for only the lexicon (i.e., morphs and/or words) needs to be stored, and everything else is computed (cf. O'Donnell 2015).



In (4), the meaning of each non-terminal constituent node is composed of the meaning of its daughter nodes, which might be complex (i.e., non-terminal) themselves: e.g., the meaning of *loved* is composed of the meaning of *love* and the past tense affix *-ed*, the meaning of *books* is derived from *book* and the plural affix *-s*, and the meaning of *loved books* is composed of the meaning of *books* and *loved*. This hierarchical constituent structure provides systematicity in grammars: constituents can be grouped into different syntactic categories whose members

compose in the same way with other linguistic material, and thus these ways of composition are definite and predictable (Pullum & Scholz 2007). If, instead of hierarchical constituent structure, we proposed a simple additive function for the derivation of (4) by positing that the meaning of *loved books* results from the combination of *loved* and *book* and then *loved book* with *-s* it would be hard to argue why the meaning of *-s* only affects the number for *book* and not *loved book* altogether and why it would not have the same meaning and/or function if used sentence initially, e.g., in “books are expensive these days”. An additive function or simple concatenation, without further semantic or syntactic structure, is sufficient for basic compositionality but not for the complex compositionality we find in natural languages.

In sum, compositional and hierarchical constituent structure is a prerequisite for a productive and productively interpretable language, and it only requires its users to learn a finite lexicon and a finite set of combinatorial rules (i.e., a grammar).

2.1.2 The cultural evolution of linguistic structure

It is a generally shared intuition that language structure is compositional because it is useful to language learning and use in communication, and studies in the field of mathematical linguistics ratify this general intuition (Pagin 2012, 2013; Yang 2016) (for a processing account on compositionality, cf. Baggio, van Lambalgen, & Hagoort 2012). But how did such linguistic structure evolve? In order to establish a causal link between the observed linguistic structure in natural languages and its functional advantages, we need to explain how the advantages of compositional structure can permeate language as a system of behaviour shared at the population level (Brighton et al. 2005; Bybee & Hopper 2001; Hall 1992; Kirby 1999).

It is probable that languages adapt over cultural time so as to maximise their learnability as long as they do not jeopardise their expressivity (Brighton et al. 2005; Christiansen & Chater 2008; K. Smith & Kirby 2012). Languages are learned from messy and relatively limited input; i.e., the learner needs to learn a grammar that generates a potentially infinite amount of data from finite data (Chomsky 1980). However, languages are robustly transmitted in spite of this learning bottleneck (Brighton et al. 2005): from limited data, language learners are able to acquire the necessary tools to generate novel expressions with minimum error. This fact about language acquisition has been used as an argument for the existence of an innate language faculty which complements the “impoverished input” and makes it possible for learners “to

know so much given so little evidence” (Chomsky 1986). The cultural evolutionary approach to language offers a complementary solution to this problem: because the poverty of the input presents a challenge to the learner, language might adapt over cultural time to maximise its learnability making its instances as generalisable as possible (Brighton et al. 2005; Zuidema 2003).

Yet, in order to be functional, it is not enough that languages are learnable, they should also be expressive (Regier et al. 2015; K. Smith & Kirby 2012). For example, a degenerate language in which all possible meanings are encoded by the same expression would be maximally learnable and could be transmitted intact but nevertheless it would also be maximally ambiguous and would not allow its users to discriminate between meanings, that is, without the incorporation of disambiguating information in the extra-linguistic context (Winters et al. 2018). On the other extreme, we have holistic languages, in which each meaning to be conveyed is expressed by a different expression—which cannot be further divided into meaningful units. This language is maximally expressive and would allow users that know it to communicate amongst themselves accurately; nevertheless, it would not allow learners to communicate about anything outside their input data and it would not survive its transmission through a learning bottleneck—unless learners are exposed to the full language and their memory resources allow them to acquire it. Only a compositional language could survive transmission as well as permit accurate and productive communication amongst its learners. Provided that a learner’s input data is sufficiently rich to allow the grammar to be deduced, the learner will acquire it alongside a lexicon and will be able to reproduce the full language as well as to produce and interpret novel expressions for novel meanings.

In the last decade, evolutionary linguists have developed experimental models to study the emergence and evolution of linguistic structure in the laboratory (for thorough reviews see Kirby et al. 2014; Scott-Phillips & Kirby 2010; Steels 2012). Kirby et al. (2008) developed an Iterated Artificial Language Learning (IALL) paradigm to specifically explore whether the pressures for learnability and expressivity previously outlined in computational and mathematical models of iterated learning (e.g. Brighton et al. 2005) would lead to similar results once idealised and rational learners were replaced with human participants. Kirby et al. (2008) showed the evolution of linguistic structure as languages were transmitted down generations of participants organised in transmission chains (Esper 1925; Mesoudi & Whiten 2008). Each

participant in a chain tried to learn an artificial language based on a set of description-meaning mappings, and during testing produced a new set of linguistic data (typed responses) which was used as the input data for the next participant in the chain. Critically, during testing participants were asked to produce description-meaning mappings that were held out during training. This data bottleneck in transmission forced participants to generalise the linguistic data available to the learner to novel descriptions for novel meanings, thus introducing a pressure for learnability (i.e., generalisability) of the linguistic data. With the implementation of this bottleneck in transmission (a pressure for learnability) in conjunction with an artificial filter to prevent ambiguous expressions to be transmitted to the next generation (a pressure for expressivity), holistic languages evolved to become compositional. Crucially, without an artificial pressure for expressivity, languages uniquely adapted to be learnable and evolved to be degenerate (see also [Perfors & Navarro 2014](#); [Silvey et al. 2015](#)).

Further studies have shown similar effects with the introduction of communicative interaction (i.e., a natural promoter of expressivity) in transmission chains, for both linguistic ([Kirby et al. 2015](#)) and graphical ([Theisen-White et al. 2011](#)) systems of communication. These studies show that sets of expressions or drawings become structured as languages are culturally transmitted through iterated learning and communicative interaction; critically, they also show that the same level of structure does not evolve from interaction alone ([Kirby et al. 2015](#); [Theisen, Oberlander, & Kirby 2010](#); [Theisen-White et al. 2011](#)). These studies thus replicate in the laboratory the results previously obtained in computational and mathematical models ([Brighton et al. 2005](#); [Kirby & Hurford 2002](#)): compositional structure results from the trade-off between learnability and expressivity pressures at play in cultural transmission.

2.1.3 The evolution of complex linguistic structure: this study

Because isomorphism (i.e., a systematic mapping between semantic and morphosyntactic structure) is a defining feature of compositional structure ([Montague 1970](#)), the complexity of the compositional languages that evolve in IALL experiments is necessarily related to the complexity of the meaning space, that is, the set of meanings participants learn and produce descriptions for. Meaning spaces utilised in the above-mentioned IALL studies are very simple. [Kirby et al. \(2008\)](#) used a meaning space that comprised 27 static pictures of coloured objects with arrows indicating motion. Each object feature (i.e., shape, color, motion arrow) varied over three pos-










object features	Shape			Colour			Motion Arrow		
feature values	square	circle	triangle	black	blue	red	horizontal	bouncing	spiraling
									

Figure 2.1: The visual stimuli used in Kirby et al. (2008) consisted of 27 pictures of coloured objects with arrows indicating motion. Each object feature (Shape, Colour and Motion Arrow) varied over three values: square, circle or triangle shape; black, blue or red colour; and arrows indicating horizontal, bouncing or spiralling motions.




	n-ere-ki	l-ere-ki	renana	□
	n-ehe-ki	l-aho-ki	r-ene-ki	○
	n-eke-ki	l-ake-ki	r-ahe-ki	△
	n-ere-plo	l-ane-plo	r-e-plo	□
	n-eho-plo	l-aho-plo	r-eho-plo	○
	n-eki-plo	l-aki-plo	r-aho-plo	△
	n-e-pilu	l-ane-pilu	r-e-pilu	□
	n-eho-pilu	l-aho-pilu	r-eho-pilu	○
	n-eki-pilu	l-aki-pilu	r-aho-pilu	△

Figure 2.2: Examples of compositional languages extracted from Kirby et al. (2008).

sible values (see Figure 2.1). In another study, Kirby et al. (2015) used a meaning space which comprised 12 pictures of patterned objects, only varying over three shapes and four patterns. Consequently, the linguistic structure that emerged in those studies is correspondingly simple, confined to referring expressions formed by concatenating simple constituents (see Figure 2.2). Although the structure is compositional so far as the meaning of the expressions is derived from the meaning of the constituent parts, there are other aspects of compositional structure found in natural languages which assist their productivity that cannot be evidenced in such simple semantic spaces. For example, non-trivial hierarchical constituent structure, which at a minimum would require the presence of complex expressions composed of complex expressions themselves, and thus of more than one of the levels of the hierarchy of constituents outlined in section 2.1.1—e.g., morphologically complex words within a complex phrase or a sentence. Moreover, as Galantucci and Garrod (2011) pointed out, IALL experiments have not shown the emergence of a type of compositionality the authors refer to as “positional”, in which the

same form takes on systematically different interpretations depending upon its position in the sequence (e.g., while ‘the dog’ is the chaser in ‘the dog chases the cat’, it is the chasee in ‘the cat chases the dog’). Altogether, these aspects of compositionality highlight the relevance of the arrangement of constituents for the derivation of meaning in natural languages; without hierarchical structure or positional compositionality, the way in which constituents are combined (i.e., grammar) is trivial because simple concatenation would suffice.

Building on previous work (Kirby et al. 2008, 2015), in the present study I aim to examine whether and how richer syntactic structure evolves through cultural transmission in the laboratory. I predict that, by introducing a more complex meaning space for speakers to learn and use, more complex linguistic structure will culturally evolve in the same way structure has been shown to evolve in previous IALL studies with simpler meaning spaces (Kirby et al. 2008, 2015). I designed a more complex meaning space that not only would posit a bigger challenge to the learner but could allow the emergence of a language which exhibits some degree of hierarchical constituent structure and argument structure. Instead of static objects I use motion events wherein the objects and motions involved are defined by two different features each (i.e., shape and number for objects; and type of motion and aspect for motions). Crucially, I include events which involve two objects with different roles (focal and anchor), and the same object can play each of the different roles thus requiring the encoding of these different roles to avoid ambiguity.

I predict that a meaning space with these characteristics will facilitate the emergence of complex nominal elements (i.e., encoding both shape and number) which can be depicted as nodes in constituent structure and thus comprise consistent syntactic categories. With the emergence of complex nominal constituents within sentences I will be able to show two levels of complex constituency and thus the minimum to show hierarchical constituent structure: morphological complex words which combine to form sentences. Moreover, the need to distinguish objects for their roles in the motion event requires the encoding of semantic roles in the argument structure either by means of morphology or word order rules. If word order rules emerge, I will be able to show the “positional” aspect of compositionality. On the other hand, if case marking systems emerge, it will also be the first time such functional morphology (i.e., case marking) emerges in IALL studies (however, see van Trijp 2012). Altogether, this more complex meaning space offers the possibility for the evolution of complex nominal elements which

are nodes in constituent structure and not mere object labels.

In addition to increasing the complexity of the meaning space, I also systematically manipulate the pressure for expressivity, following previous work in assuming that compositional structure will arise from a trade-off between pressures for expressivity and learnability. I do so through a series of experiments in which I model a learnability pressure in language transmission and vary the nature of a competing pressure for expressivity in language production: in Experiment 1 I introduce an artificial pressure against ambiguity (Carr, Smith, Cornish, & Kirby 2016; Kirby et al. 2008; Verhoef 2012) and in Experiment 2 I implement communicative interaction, a more naturalistic pressure for expressivity (Kanwal et al. 2017; Kirby et al. 2015; Winters et al. 2015). In a further Experiment 3, I combine transmission with both communication and an artificial pressure against ambiguity.

2.2 Experiment 1: transmission and artificial pressure against ambiguity

2.2.1 Method

The experiment utilises an Iterated Artificial Language Learning paradigm (Kirby et al. 2008, 2015). In overview, each participant in a transmission chain tries to learn an artificial language based on some linguistic data, and then during testing produces a new set of linguistic data which will be the input data for the next learner in the chain. I ran four transmission chains as per Kirby et al. (2008), each containing eight generations of participants. Following Verhoef (2012), I implemented a strict artificial pressure against ambiguity during testing to avoid the evolution of degenerate languages.

2.2.1.1 Participants

Thirty-two participants were recruited to participate in an artificial language learning study through the University of Edinburgh Careers Service database of student and graduate employment. All participants were native speakers of English (mean age 22 years, age range 18–42). Participants received a payment of £9. The experiment was conducted in accordance with the ethics procedures of Linguistics and English Language, The University of Edinburgh.

2.2.1.2 Stimuli

Participants were asked to learn, and then reproduce, an artificial language which provided descriptions for scenes of motion events. Motion events were represented using videos, descriptions were presented as text labels above the videos. I created 80 animated scenes to represent the motion events. Each scene was five seconds long and featured a scene with one or more objects performing a motion; in some scenes this motion took place on a blank screen, in others the motion was relative to another object or set of objects.

More precisely, each scene featured a *focal* object or objects and, optionally, an *anchor* object or objects. There were two types of objects, squares and circles; each object could appear singly or as part of a group of multiple (i.e., nine) objects of the same shape (e.g., a group of nine circles, a group of nine squares). The focal object(s) in each scene performed one of two possible motions: sliding across the screen, or bouncing across the screen. That movement could occur once (resulting in a terminated motion event) or be continuously repeated for the entire duration of the scene (producing an ongoing motion event). If the scene featured anchor objects, the focal objects were initially on the opposite side of the screen from the anchor objects and moved towards the anchor objects; in scenes lacking anchor objects, the focal objects simply started on one side of the screen and moved to the other. The initial position of the focal objects (left or right side) was randomised on each presentation of each scene.

More formally, each motion event differed on five binary features: Shape of focal object, Number of focal objects, Motion, Aspect (terminated vs. ongoing), and Anchoring (whether the event comprises an anchor object or not). Events with anchor objects(s) differed along two further binary features: Shape of the anchor object and Number of the anchor object—which contained the same features as Shape and Number of focal objects. This yields the full set of 80 possible motion events (16 events lacking anchor objects, 64 featuring anchor objects). Figure 2.3 provides a visualisation of the meaning features and values described.

This set of 80 stimuli defines the meaning space for which participants will be asked to produce descriptions in artificial languages. If compositional languages evolve, the presence of scenes with both focal and anchor objects, which both need to be defined by two meaning features at a minimum (i.e Shape and Number), will facilitate the emergence of mappings to complex nominal constituents (i.e., including morphs encoding Shape and Number) that

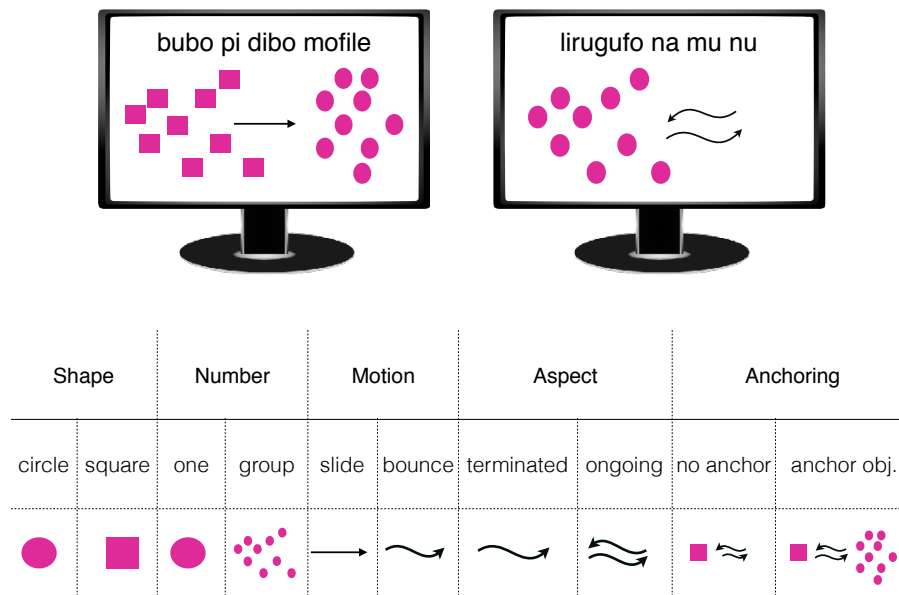


Figure 2.3: *Features and values of the stimuli scenes and example stimuli.* The reference space in this experiment consists of events which are composed of 5–7 different features (depending on the presence or absence of an anchor object respectively), each comprising two possible meaning values. In the header of the table (bottom) I show the different features (collapsing Shape and Number in focal and anchor objects); on the rows I list the different values that constitute each meaning feature and the correspondent illustrations. Above the table, two examples of stimuli (with the motion represented with arrows) as it would appear on the screen to participants during the learning phase: they would see a motion picture with the corresponding label on top of the scene. The screen on the left shows a group of squares sliding towards a group of circles. The screen on the right shows a group of circles bouncing back and forth (without any anchor object).

will be differentiable from the other elements within descriptions (i.e., those referring to the Motion and Aspect features). Moreover, the fact that objects can appear in different roles within motion events will require an expressive language to encode whether objects are focal or anchor objects in the motion event. This meaning space thus affords the evolution of complex nominal elements that can form a node in constituent structure as well as the use of word order rules and/or morphology to encode semantic roles.

2.2.1.3 Initial languages

The initial languages to be learned by the first participant in each chain were a set of randomly generated holistic strings of lower-case letters, possibly including spaces. For each initial language, I generated 80 unique strings: each string consisted of 2–8 CV syllables, divided by

spaces into 1–8 chunks (the number of chunks was randomly selected). These 80 strings were then paired randomly with the set of motion event scenes to create 80 scene-description pairs. I generated a separate initial language for each initial participant in each chain in order to eliminate any specific biases that might be imposed by the initial language.

2.2.1.4 Procedure

Training and testing regime Participants were asked to learn an artificial language made up of written labels for visual stimuli which they would be trained and tested on. They carried out the experiment at a computer terminal in isolated individual booths. All responses were entered using the keyboard. Participants received written and verbal instructions before starting the experiment, and on screen at the start. The experiment was divided into two phases: a training phase and a testing phase.

During training, participants were taught a subset of 44 scene-description pairs from the total 80 pairs (randomly selected but always containing 3/4 of the non-anchored events and 1/2 of the anchored events). Each pair was presented three times in randomised order, yielding a total of 132 training trials, for a training phase duration of approximately 30 minutes. In each training trial, the description was shown in isolation for 1 second, then the associated scene was shown, accompanied by the signal, for 5 seconds. After each presentation, participants underwent one of two recall tests: retyping (50% of the time) or scene discrimination (other 50% of the time) (randomly chosen). In the retyping recall test, participants were presented with the motion event they had just seen, and were asked to retype its description. In the discrimination recall test participants were presented with two scenes side by side in randomised position, one of which they had just seen and the other selected randomly from the total set of 80 scenes; they were then asked to identify by button-press which of the two motion events matched the one they had just seen. These recall tests were intended to ensure that participants attended both to the training descriptions and their associated scenes.

During testing, participants were asked to describe all scenes twice in randomised order, yielding a total of 160 testing trials (approximate duration 40 min). Note that, since the participants were trained on 44 scenes of motion events but tested on all 80, this meant they were tested on events they had not been trained on: a post-test oral questionnaire revealed that none of the participants realised that they had been tested on untrained scenes. On each testing

trial, the participant was presented with a scene for 5s (this time without a description), and then asked to type its corresponding description in the artificial language. Participants were told that the computer could interrupt them to ask for another description if they had already typed that same exact string to describe a different motion event; participants were therefore prompted to produce a different description whenever they entered a string which they had already used to describe a different scene during testing. This explicit demand for unique descriptions is intended to introduce a pressure against ambiguity to prevent the language from collapsing to a maximally-ambiguous single description (see Verhoef 2012).

Transmission Participants were organised into independent transmission chains (for review see Mesoudi & Whiten 2008), such that the language (set of scene-description pairs) produced by a participant at generation g is used as the training language for another participant at generation $g+1$ in that chain of transmission. Languages were formed by the set of descriptions participants last produced for each meaning. The initial participant in each chain, the first generation, is trained on a random target language, generated as described in section 2.2.1.3.

As mentioned in the training and testing procedures, I imposed a *bottleneck* on transmission. A language (either constructed with random strings or produced by a participant) consists of a set of 80 scene-description pairs. During transmission, I divide this into two subsets, a trained set (44 scenes, selected randomly) and an untrained set (the remaining 36 scenes) as described in section 2.2.1.4. This sub-setting procedure is implemented at each generation in a chain: the first participant is trained on a subset of the initial target language, subsequent participants are trained on a subset of the previous participant's output language. Participants were not informed of the source of the artificial language (i.e., that it was produced by another participant) until after completing the experiment.

2.2.2 Measures

2.2.2.1 Compositional structure: isomorphism between semantic and syntactic structure

Following Kirby et al. (2008, 2015), I quantify compositional structure as the z-score of the Mantel Test between description similarities and scene similarities. Description similarity is calculated using normalised Levenshtein distance (Levenshtein 1966), which is the number of

characters that need to be changed, inserted or deleted to transform a description into another divided by the length of the longest description (such that the maximum distance is 1). Scene similarity is calculated using Hamming distance (Hamming 1950), which is given by the number of feature values that are different between two scenes. Thus to quantify structure in a language, I first calculate the correlation coefficient between all pairs of edit distances in the set of descriptions and all pairs of edit distances of the corresponding scenes. This veridical coefficient gives us an indication of the extent to which similar meanings are associated with similar signals, as would be expected in a compositional language. I then calculate how likely the veridical coefficient between the two distance matrices is to appear by chance, using the Monte Carlo method of random sampling to produce a distribution of coefficients. At each sample I randomise the associations between meanings and signals and re-calculate the correlation. I ran 10,000 samples, and from the distribution obtained, I extract the z-score for the veridical coefficient. If the z-score is greater than 1.645 (one-tailed)¹, I conclude that the veridical coefficient is unlikely to arise by chance ($p < 0.05$). High z-scores thus indicate a higher degree of compositionality on this measure.

2.2.2.2 Reference

In order to minimise the influence of human biases in the linguistic analysis of the descriptions produced by participants, I extract form-meaning mappings automatically. I identify the referents of the lexical items² in the miniature artificial languages by calculating the association strength between lexical items and meaning feature values of scenes. I use Kendall's Tau-b rank correlation coefficient (Kendall 1938, 1945), which allows me to measure the strength and direction of the correlation between occurrences of a given lexical item and those of a given meaning feature value. Values of Tau-b range from -1 to +1, indicating 100% negative or positive association respectively. A value of 0 indicates the absence of association. Thus the more a lexical item co-occurs with a specific feature value, the higher above 0 the Tau-b is. Thus the more distant Tau-b is from 0, the stronger the referential association.

¹I use one-tail critical z-score values because I do not predict large negative z-scores; i.e., I do not expect significant negative correlations between description-similarities and scene-similarities. Moreover, as seen in the results section later on, z-scores obtained are very large and thus the use of one-tail instead of two-tail critical values does not alter the results.

²Lexical items are each of the strings separated by spaces (introduced by the participants themselves) within a description.

Before calculating Tau-b coefficients, I run a diagnostic test to provide a more robust threshold for the significance of the dependence between lexical items and meaning feature values: I compute the mutual information of all pairs of lexical items and meaning feature values in a language. For each pair, I use the Monte Carlo method of random sampling to calculate how likely this veridical mutual information is to appear by chance. At each sample I randomise the mapping between scenes and descriptions and re-calculate the mutual information between the lexical item and the feature value. I run 10,000 samples, and from the resulting distribution, we calculate the probability to obtain by chance a mutual information equal or higher than the veridical; only if $p < 0.05$ I conclude a significant mutual dependency between a given pair (i.e., between a lexical item and meaning feature value) and proceed to calculate its Tau-b coefficient. Mutual information provides us with non-spurious correlations but not with a normalised value for the strength of the correlation or its direction (either positive or negative). Both strength and direction are obtained with Tau-b.

2.2.2.3 Nominal syntactic categories

The emergence of complex nominal constituents (encoding both shape and number meaning features) is crucial to our study because they will provide the evidence required for hierarchical structure as well as for positional compositionality. In order to conclude the emergence of systematic complex nominal constituents I need to demonstrate that constituents associated with shape meaning features appear adjacent to number meaning features and crucially, that they constitute a syntactic category. The syntactic category of a given grammatical unit can be inferred from the distributions in which it appears within sentences. To assess whether constituents which contain morphs that refer to the shape-objects in the scenes³ form a syntactic category in the miniature artificial languages, I calculate the distance between their distribution in descriptions, i.e., their distributional distance. I quantify the distributional properties of these constituents in a language (which I call nominals henceforth) as their set of backward transitional probabilities (BTP) (following McCauley & Christiansen 2011; Perruchet & Desautly 2008), i.e., the probabilities that each nominal has of being preceded by each of the lexical items in a language's lexicon. I then use the Jensen-Shannon distance metric (JSD) to measure the distances between all pairs of BTP distributions. The lower the distance between each pair

³Calculated as described in section 2.2.2.2

of distributions, the more similar they are. The average JSD between all pairs then gives us an indication of the distributional similarity between all nominals. I then calculate how likely this veridical average JSD is to appear by chance using the Monte Carlo method of random sampling. At each sample, I randomly select a set of lexical tokens of the same cardinality of the set of nominals in the language and calculate the average JSD between all pairs of BTP distributions. I run 10,000 samples, and from the distribution obtained, I extract the z-score for the veridical average JSD. Low z-scores indicate short distances between BTP distributions of nominals and specifically, z-scores below -1.645 ($p < 0.05$, one tailed)⁴ suggest that nominals within a language share similar distributions and thus constitute a syntactic category.

2.2.2.4 Order of nominal arguments

The same object or group of objects can appear both in focal and anchor roles within the motion events in the stimuli. There are different ways in which a language describing the stimuli could mark nominal arguments for the semantic roles they perform. Whether an object is in the role of focus or anchor could be morphologically encoded (e.g., via affixation, suppletive forms or functional particles), but it can also be cued by the order in which nominals appear within a sentence. For example, nominals appearing in first and second position in a sentence can be systematically assigned focal and anchor semantic roles respectively; in this case, the position in which nominals appear could determine their meaning. I assess the systematicity of the order of nominal arguments by calculating the Shannon entropy of the different orders in a language. Entropy measures how variable the order of nominal arguments is between sentences in a language. In an unambiguous compositional language there are only two possible nominal orders, either focal arguments precede anchor arguments (focal-anchor orders) or vice versa (anchor-focal orders). Nevertheless, as linguistic structure emerges, often the order of nominals will be undefined: participants will produce underspecified descriptions either because they might not have the lexicon to refer to objects and/or the number they appear in, or because they simply do not use the lexicon consistently. In order to measure the entropy of the system of nominal argument orders, I first exclude from the language those descriptions that

⁴I use one-tail critical values because I do not predict large positive values. Large positive z-scores could only be obtained if the distribution between nominals were more dissimilar than obtained by chance. Given the conservativeness of the random sampling method (i.e., selection from tokens and not types, and descriptions are kept intact at each randomisation), I do not expect to obtain dissimilarity scores significantly distant from the mean of the random sample.

refer to motion events which do not have an anchor object (16 descriptions) and those whose focal and anchor object are the same (a further 16 descriptions)—as the order of nominals there is not informative. I then calculate the frequency of focal-anchor, anchor-focal and undefined orders within the reduced language of 48 descriptions. Undefined orders introduce randomness into the system; in order to implement such randomness in the entropy measure, I split the frequency of undefined orders between focal-anchor and anchor-focal patterns equally. I then calculate the entropy of the resulting vector of frequencies. A language without any defined pattern of the order of nominal arguments obtains the maximum possible entropy of 1 bit (i.e., a language with 50% anchor-focal and 50% focal-anchor orders, or a language with all undefined patterns), and a language with a consistent order of nominal arguments would result in the minimum possible entropy, i.e., 0 bits (i.e., a language with 100% anchor-focal or 100% focal-anchor orders).

2.2.3 Analysis and results

2.2.3.1 Languages

Before presenting further quantitative analysis of linguistic structure I first introduce the languages which evolved in the experiment. I analyse in detail the morphosyntactic structure of an example language obtained amongst the four chains. Later, I present an overview of the structure of all four languages.

For each language produced in Experiment 1, I extracted a matrix of associations between lexical items and their referents in the scenes as explained in section 2.2.2.2. With this matrix I was able to automatically gloss the meanings of the descriptions provided in the experiment and analyse their structure minimising any potentially biased interpretation of the semantics of a language. The induced dictionary for the final language A3 (where the A stands for Artificial plus transmission and the 3 indicates the chain number) is shown as a matrix of associations in Figure 2.4 as an example. Examples of descriptions in the same language A3 with the corresponding glosses are provided in (5). In these descriptions, I observe that lexical items associated with shape-objects, which I call nominals, precede the lexical items associated with motion or aspect features in the scenes, which I refer to as verbal elements. Moreover, Language A3 uses word order as a morphosyntactic cue to interpret the different semantic roles

of the nominal arguments in a sentence; the roles of focal and anchor object are consistently assigned to the first and second position in a description respectively. The correct interpretation of the semantic roles of each of the nominal arguments is crucial, as the same nominals can refer to focal and anchor objects. Language A3 also makes use of redundant functional markers of semantic roles and anchoring (i.e., the presence of an anchor object in the event) (see sentences 5a–b): one marker, *pifli*, systematically follows focal nominal arguments, and another marker, *trink/-i san/-s*, follows anchor nominal arguments. In addition, the form of the latter is conditioned by number of the nominals: if one or more nominals are marked as plural, these anchoring markers will appear as *trinki sans* rather than *trink san* (see 5a-b).

- (5) a. rons pifli mons trinki sans hula bu
 square.group ANC⁵ circle.group ANC.group slide terminated
 ‘A group of squares slid towards a group of circles’
- b. mon pifli ron trink san hula
 square ANC circle ANC.one slide
 ‘A circle slides towards a square back and forth’
- c. mons hulai ai
 circle.group bounce terminated
 ‘A group of circles bounced’

Language A3 comprises two main categories: a nominal category that consists of complex constituents formed by morphs associated with the Shape and Number features (the latter always suffixed to the former), and a verbal category formed by morphs associated with Motion, followed by a marker of Aspect in terminated events. Nominals are thus morphological complex lexical constituents with strict internal structure. Any linear word order rule has to respect the integrity of the nominals and cannot break them up, i.e., number will not be realised if it is not as a suffix in nominals.

Tables 2.1 and 2.2 below display the different lexical items under the nominal (Table 2.1) and verbal categories evolved in all languages. Languages A1–3 are extracted from the final generations and language A4, from the penultimate generation—the participant in the last

⁵ANC stands for anchoring marker, particles that appear only with events that contain both a focal object and an anchor object.

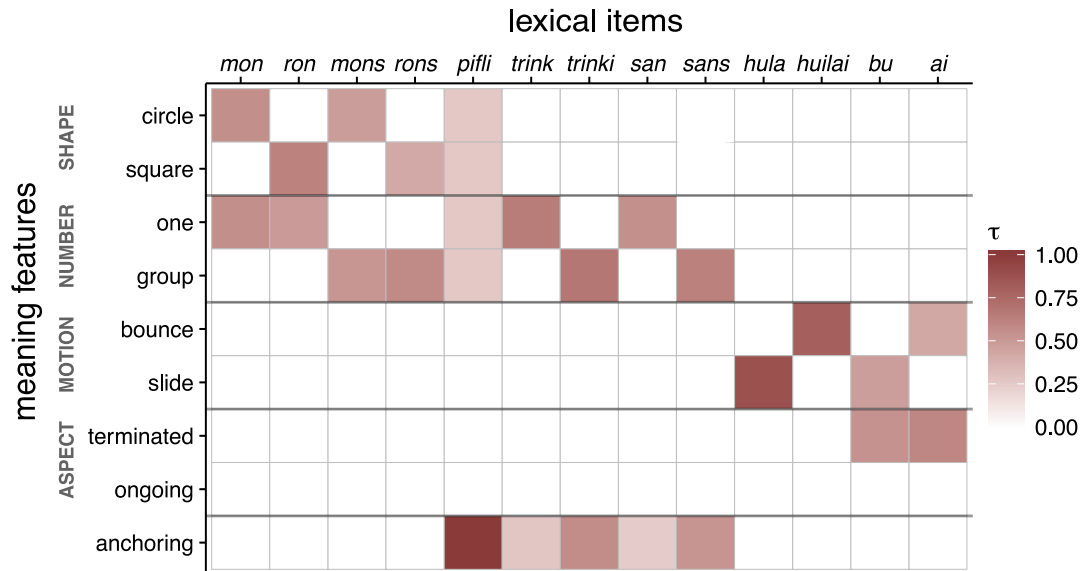


Figure 2.4: The above heatmap illustrates the different semantic categories of lexical items (x axis) found in language A3 (chain 3) in relation to the meaning features and values they refer to (y axis). The heatmap scale represents the strength of the positive association between lexical items and meaning values (Tau-b coefficient, see section 2.2.2). We can distinguish three salient patterns in A3’s lexicon which correspond to three different categories that I will call nominal, functional and verbal elements. Moving from left to right along the x axis in Figure 2.4 we find: lexical items associated with Shape (*circle* or *square*) and Number (*one* or *group*) which form a nominal category, a set of lexical items associated with Anchoring (presence or absence of an anchor object) as well as Number which correspond to the only functional category, and lexical items associated with Motion (*slide* or *bounce*) and/or Aspect (*terminated* or *ongoing*), which constitute a verbal category. In the nominal category we have *mon/-s* and *ron/-s*, which are the only items that refer to the shapes in the scenes. The affix *-s* acts as a plural marker and its absence marks singularity. Verbal elements are *huilai*, *hula* and *ai* and *bu*. Both *huilai* and *hula* are free morphs associated with Motion alone, and *ai* and *bu* act as their respective aspect morphemes (i.e., although separated by spaces, they cannot stand on their own), whose presence marks the events as *terminated*.

		Lang A1	Lang A2	Lang A3	Lang A4	(if focal object & Motion = slide)
circle	one	sunyan	piona /pijone	mon	cica	(cicaa)
	group	sanyan	piondra /e	mons	lumuse	
square	one	vunyan	fiona	ron	demi	(dmei)
	group	vanyan	fiondra /e	rons	demi-toda/tofa/fora	

Table 2.1: Nominals in the final languages: at generation 8 for languages A1–A3 and generation 7 for language A4. The elements in bold signal category markers, recurrent patterns across members of a lexical category.

		Lang A1	Lang A2	Lang A3	Lang A4
slide	ongoing	F tolo A (vale)	watashe, zu, yu, mebe	F A hula	F fumuse A
	terminated	F vero/velo A (re/te)	watashe, zu, yu, mebe	F A hula bu	F sahime A
bounce	ongoing	F galamete A (vale)	watashe, zu, yu, mebe	F A huilai	F fumuse A
	terminated	F vero/velo/galamete A te/re	watashe, zu, yu, mebe	F A huilai ai	F sahime A

Table 2.2: Verbal elements in the final languages: at generation 8 for languages A1–A3 and generation 7 for language A4. Elements in brackets are optional. The most frequent position of anchor nominal arguments is represented by *A*, and that of focal nominal arguments, by *F*.

generation of chain 4 failed to learn the lexical items of the input language causing a drastic decrease in the language’s structure (see section 2.2.3.2 to follow). Sentences mostly comply with the structures described in the two tables and therefore can be reconstructed from them. The order of nominal arguments is mainly fixed in A1, A3 and A4: focal nominal arguments precede anchors.

All languages encode Shape and Number within nominals. Moreover, we observe further sublexical structure in the morphs encoding Shape within nominals: most languages (A1–A3) contain a nominal category marker, signalled in bold in Table 2.1. Number is almost exclusively marked via affixation (simulfixation in A1 and suffixation in the rest) to the exception of the suppletive forms found in language A4 to mark the plurality of the *circle* objects (see Table 2.1).

Whilst the encoding of Shape and Number is fairly systematic in the final generations across chains, the encoding of Motion and Aspect is not entirely or at all established in half of the languages (e.g., A1 and A2 respectively). Moreover, it is only in the last two generations that Motion starts to be encoded in language A4 but it is not entirely systematic either⁶.

⁶This late and sudden encoding of a previously underspecified meaning feature suggests that the participants

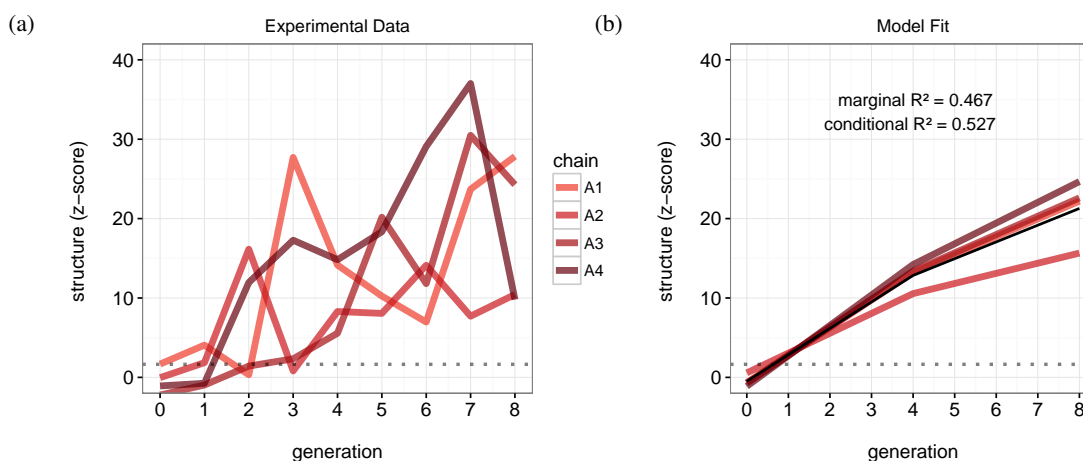


Figure 2.5: (a) Linguistic structure over generations for each of the four transmission chains. Linguistic structure increases as languages are transmitted through generations of learners. (b) Fitted values from the mixed-effects regression Model 1 for the four transmission chains and their average (in black). Coloured lines represent the random slopes estimates (for generation) depending on random intercepts (individual chains), whereas their average in black represents the fixed effects estimates. In both plots, the dotted horizontal line represents the chance level (z-score 95%CI = ± 1.645 , one-tailed).

2.2.3.2 Compositional structure

I hypothesised that, by introducing a more complex meaning space for speakers to learn and communicate about, more complex linguistic structure would culturally evolve in the same way structure has been shown to evolve in previous IALL studies with simpler meaning spaces (Kirby et al. 2008, 2015). As discussed in section 2.2.3.1, linguistic structure indeed emerges to convey this complex meaning space through cultural evolution. Languages shift from holistic to compositional systems. Figure 2.5a shows the structure scores obtained in the experimental data (see measures in section 2.2.2.1). We observe that structure gradually increases as languages are transmitted through generations of participants; all languages are significantly structured from generation 4 onwards (chance level is represented by a dotted line in Figure 2.5a).

I used R (R Core Team 2000) and the package *lme4* (Bates, Mächler, Bolker, & Walker 2015) to perform a segmented linear mixed-effects model (SLMM) (as described in Baayen 2008) to explore the effect of generation on linguistic structure (measured as explained in section 2.2.2.1). For ease of reference I will call this Model 1. I ran a SLMM because it allows to easily quantify an abrupt change of the response function of a varying influential

might be employing a conscious strategy to increase expressivity.

factor. I expect the influence of generation to be more distinct in the first generations and significantly less so in the latter generations as languages become structured. Unlike other types of growth curve modelling (e.g. Growth Curve Analysis), SLMM allows me to identify a specific point of change in an otherwise linear relation between generation and the dependent variable (structure), and most importantly, the direct effect of that point of change. In Model 1, Generation is partitioned into two intervals with one breakpoint at generation 4, and a separate line segment is fit to each interval⁷. In order to extract the most adequate breakpoint I followed the standard procedure described in (Baayen 2008, pp. 238–239). I first fitted a series of models, one for each possible breakpoint in the range of generations, including breakpoints at generation 0 and 8, which equate to no breakpoint⁸. I then selected the breakpoint of the model with the lowest deviance⁹, which was at generation 4. As fixed effects, I entered Generation (shifted) with Indicator nested, which equates to the effect of Generation and an interaction between Generation and Indicator. As random effects, I introduced intercepts for Chain as well as by-Chain slopes for the effect of Generation. The overall model fit is $R^2_{\text{marginal}} = 0.467$ and $R^2_{\text{conditional}} = 0.527$; Figure 2.5b shows the predicted values based on the fixed and random parameter estimates obtained. I found a significant¹⁰ effect of Generation ($\beta = 3.332, SE = 1.08, p = 0.005$), suggesting that structure increased as languages were transmitted through generations of learners. There was no significant interaction between Generation and Indicator ($\beta = -1.222, SE = 1.831, p = 0.509$), indicating that structure did not increase more in the first four generations than in the later generations. These results suggest that structure was still increasing in the last generations as much as in the first generations.

⁷In order to introduce a breakpoint at generation 4, I first shifted the value of Generation so the intercept at 0 is in generation 4. I then introduce an Indicator variable that specifies whether or not each of the shifted values is greater than 0.

⁸It is worth noting that this automatic procedure developed in Baayen (2008) yields a higher-than-nominal Type-I error rate of finding non-linearity. In order to check that multiple comparisons were not too problematic in the models presented in this chapter I followed the simulation-based approach described in Vanhove (2014) to calibrate the p-values. In short, I simulated 10,000 datasets based on the experimental data, looped through each of them to determine the individual best-fitting segmented (fixed-effects only) models, and saved the associated p-value the effect of the Indicator variable. To calibrate the observed p-value against the distribution of p-values obtained from the simulated models, I just looked up the proportion of p-values generated under the null hypothesis equal to or lower than the observed p-value. If this proportion was lower than 0.05, I kept the segmented model over the simpler linear model.

⁹Note that the lowest deviance was also extracted across the models comprising the data from experiments 2 and 3 in this chapter; we include it in this simpler model even though it is not significant to assure consistency throughout.

¹⁰As in all models to follow, p-values were calculated using *lmerTest* (Kuznetsova, Brockhoff, & Christensen 2014). The library *lmerTest* calculates p-values of fixed effects from F statistics based on Satterthwaite's approximation for denominator degrees of freedom, and it tests random effects using likelihood ratio.

	gen 1	gen 2	gen 3	gen 4	gen 5	gen 6	gen 7	gen 8
chain A1	na	na	1	1	1	1	1	1
chain A2	na	1	na	1	1	0.9	na	1
chain A3	na	na	1	1	0.96	0.96	0.93	0.76
chain A4	na	0.97	1	1	1	1	1	1

Table 2.3: Proportion of adjacent Shape and Number morphology for a specific object, either focal or anchor. Shape and Number are always encoded adjacent to each other in most languages from generation 2 onwards. The only case where we observe a notably lower proportion of adjacency is in the final generation of chain A3, where on top of Shape and Number morphology, case-like markers evolve which can agree in number with distant objects and not the immediate one.

2.2.3.3 Hierarchical constituent structure: the emergence of complex nominal constituents within sentences

I hypothesised that complex constituents would evolve given the affordance of the meaning space participants were asked to describe. In particular, I hypothesised that morphologically complex nominals which constitute a node in constituent structure could emerge, comprising at least morphs that refer to Shape and Number meaning features. Table 2.3 shows the relative frequency in which morphs referring to the Shape and Number features of a specific object (either focal or anchor) appear adjacent to each other within descriptions. We find that they consistently appear adjacent to each other in most languages once morphs to encode shape and number evolve, i.e., from generation 2 onwards. Number morphs across languages are bound to Shape morphs via affixation, specifically either by suffixation or simulfixation. The only case where we find a notably lower proportion of adjacency is in the final generation of chain A3, where, as described in section 2.2.3.1, on top of number morphology adjacent to Shape, long-distance number agreement evolves (encoded in redundant functional markers).

I now turn to test whether these complex constituents can be syntactically categorised as nodes in the structure of yet more complex linguistic expressions, i.e., sentences. Syntactic categories are formed by constituents which arrange with other linguistic material in a similar way and thus share the same distributions in a sentence. I assessed the significance of the distributional distance between nominal constituents as explained in section 2.2.2.3. Figure 2.6a shows the z-scores of the average distance between the distribution of nominals in the descriptions of language; z-scores below -1.645 indicate that the distributional similarity between nominals is unlikely to result by chance. We observe the emergence of nominal syntactic cat-

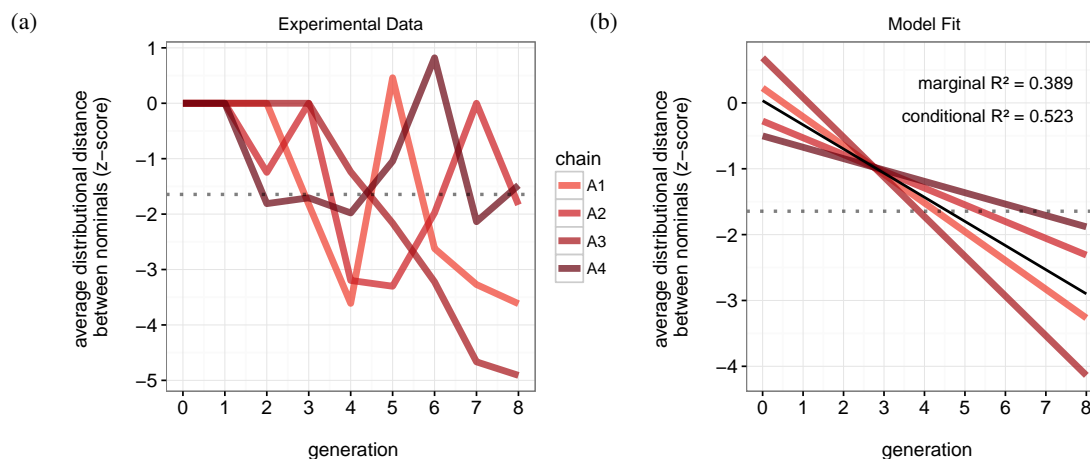


Figure 2.6: (a) Distributional distance between a language’s nominals through generations for each of the four chains. The dotted line represents the chance level (z -score 95%CI = ± 1.645 , one-tailed); z -scores below it indicate that the distributional similarity between nominals is unlikely to arise by chance. Distributional distance between nominals decreases with generation as languages become more structured, suggesting the emergence of a nominal syntactic category. Nonetheless, only two languages stay consistently below chance from generation 6 onwards and only three end up below chance at generation 8. (b) Fitted values from the mixed-effects regression Model 2 for the four transmission chains and their average (in black). Coloured lines represent the random slopes estimates (for generation) depending on random intercepts (individual chains), whereas their average in black represents the fixed effects estimates.

egories across chains, i.e., the different nominals share similar distributions in the descriptions of the language. Nevertheless, only two languages (chains A1 and A3) stay consistently below chance from generation 6 onwards, the other two languages (chains A2 and A4) are less stable and are distributed around chance level ($z = -1.645$) by the final generation. Note that given the conservative nature of the analysis of distributional distance, which is carried out on unannotated languages, if we obtain z -scores significantly below chance we should assume that the order of verbal elements in relation to nominals is mostly fixed. I cannot infer anything about the order of nominals as this measure is blind to semantic roles (i.e., whether nominals refer to focal or anchor objects); I discuss the order of nominals within descriptions in relation to their semantic roles in the following section 2.2.3.4.

I also performed a linear mixed-effects model, which I will call Model 2, to explore the relationship between the distributional distance of nominals and generation. I did not run a segmented linear mixed model because the absence of a breakpoint constituted the best fit to the experimental data. I entered Generation as the only fixed effect (centred). As random effects, I introduced an intercept for Chain and a by-Chain random slope for the effect of

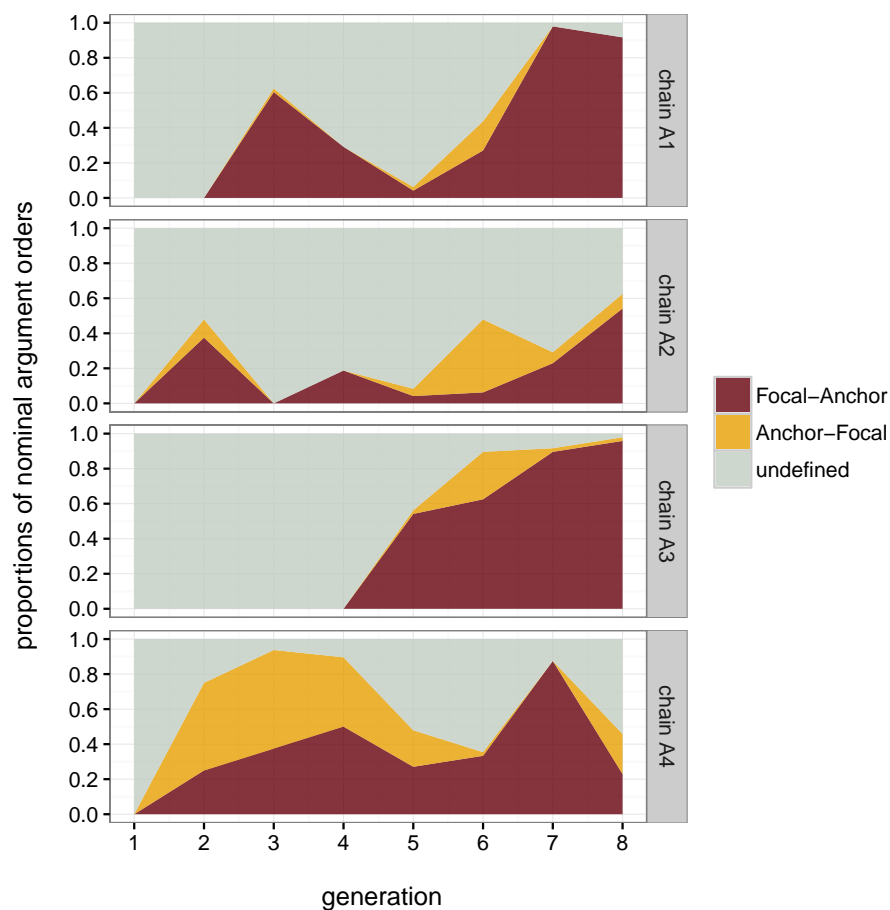


Figure 2.7: Proportion of word order types for nominal arguments by chain and generation. The proportion of undefined word orders decreases as languages become more systematic and in chains A1 and A3 we observe the evolution of a fixed Focal-Anchor word order towards the final languages. In chain A4, word order becomes systematically Anchor-Focal at generation 7 but de-systematises in the next generation as the participant fails to learn the vocabulary of the language. In the remaining chain A2, word order rules never evolve.

Generation. The overall model fit was $R^2_{\text{marginal}} = 0.389$ and $R^2_{\text{conditional}} = 0.523$. Figure 2.6b shows the fitted values of Model 2 for fixed and random effects. Results showed a significant effect of Generation ($\beta = -0.367, SE = 0.120, p = 0.036$), suggesting that the distributions in which complex nominals appear do become more similar as languages are transmitted through generations of participants.

2.2.3.4 Word order rules for nominal arguments

Languages have various ways of encoding the semantic roles of the arguments in a sentence. Semantic roles can be encoded morphologically or can be cued by the position they occupy in a sentence. In section 2.2.3.1 we observed that all nominals could occupy both focal and

anchor semantic roles, and only one of them (i.e., A3) developed a morphological marker for the different semantic roles, which followed nominals. The stacked area graphs in Figure 2.7 show the proportions of focal-anchor, anchor-focal and undefined orders at each generation for each of the four transmission chains in the experiment. As explained in section 2.2.2.4, these relative frequencies are taken from the set of 48 descriptions that refer to scenes which include two objects (one in focal and another one in anchor roles) whose shape and/or number differ. A visual inspection of Figure 2.7 suggests that the proportion of undefined order decreases as languages are transmitted through generations. Moreover, we observe the evolution of a fixed order of nominal arguments in at least three out of the four chains. In languages A1 and A3, the order of nominal arguments is mostly fixed in the last two generations; the proportion of focal-anchor order is $\geq 90\%$. I thus conclude word order rules emerged to convey semantic relations in A1 and A3, where the same nominal can function as a focal argument or an anchor argument depending on its position; in both languages, focal arguments precede anchors (for examples, see 2.2.3.1). We observe the same tendency in language A4, where a focal-anchor nominal order becomes increasingly established and constitutes the 87% of nominals orders at generation 7; nevertheless, generation 8 produced a much more unstructured language (see Figure 2.5 in section 2.2.3.2) resulting also in an increase of undefined order of nominal arguments. However, undefined orders in language A2 remain the norm throughout generations and only at the final generation 8 we observe an increase of the focal-anchor order, but only to reach 54%.

I ran a linear mixed effects model, which I will call Model 3, to test the effect of generation on the variability of nominal argument orders in a language —measured by the entropy of the system of orders as described in section 2.2.2.4. I entered Generation (centred) into the model as the only fixed effect; as random effects I enter intercepts for Chain¹¹. Figure 2.8 shows the nominal order variability scores of the experimental data (Figure 2.8a) as well as the fitted values of Model 3 for fixed and random effects (Figure 2.8b). The overall model fit was $R^2_{\text{marginal}} = 0.349$ and $R^2_{\text{conditional}} = 0.367$. Results show a significant effect of Generation ($\beta = -0.06, SE = 0.017, p < 0.001$), suggesting that the order of nominal arguments becomes more consistent as languages are transmitted through generations of learners. Therefore, along

¹¹This was the maximum random effects structure allowed without convergence warnings. Moreover, the model with the inclusion of by-Chain random slopes for the effect of Generation (maximal random effects structure) was not significantly better ($\chi^2(2) = 0.597, p = 0.742$).

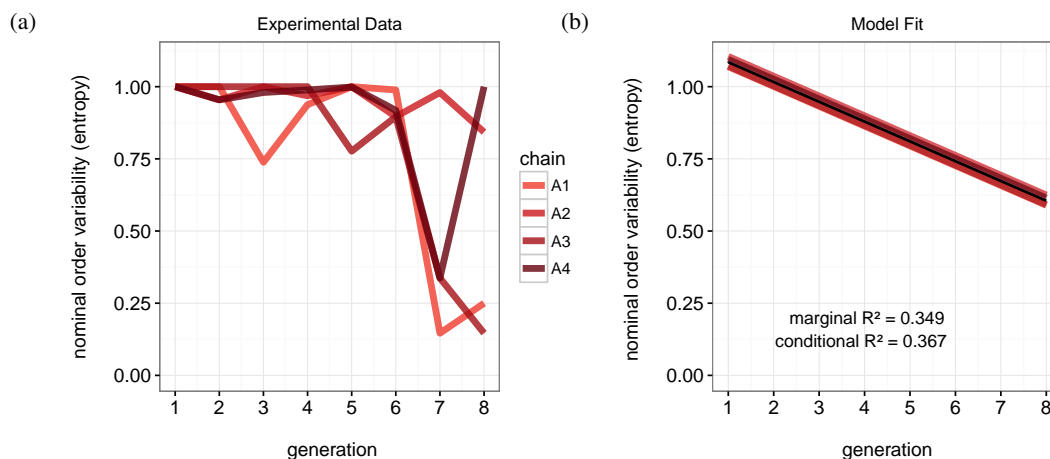


Figure 2.8: (a) Variability of nominal argument orders by generation and chain. (b) Fitted values from the mixed-effects regression Model 3 for the four transmission chains and their average (in black). Coloured lines represent the values estimated for the different random intercepts (i.e., each individual chain), whereas their average in black represents the fixed effects estimates.

with an increase of overall structure, the order of nominal arguments becomes more consistent suggesting the emergence of what Galantucci and Garrod (2011) called “positional” compositionality: the same exact nominal constituent can acquire different semantic roles (either focal or anchor) depending on its position.

2.2.4 Discussion

In Experiment 1 I examined whether complex compositional structure would evolve in the same way basic compositionality has been shown to evolve in previous IALL studies by introducing a more complex meaning space for speakers to learn and communicate about (Kirby et al. 2008, 2015). I designed a complex meaning space that could allow the emergence of a more complex structure that mirrors the one found in natural languages more closely than previous IALL studies have shown. First of all, I wanted to test whether linguistic structure would evolve with such a complex meaning space and thus whether complexity would limit the IALL experimental model in the evolution of compositional structure (cf. Carr et al. 2016 for the evolution of systematic but not compositional structure in an open-ended continuous meaning space). However, most importantly, I wanted to provide evidence for two properties of compositional structure found in natural languages that had not yet been shown in IALL studies: hierarchical constituent structure (i.e., complex constituents are built from further complex constituents),

and argument structure whose semantic roles are marked via word order rules evidencing “positional” compositionality (i.e., the same constituent takes different semantic roles depending on its position in a sentence).

Compositional structure evolved from holistic languages as they were transmitted through generations of participants. Languages developed morphology to match existing feature values in the meaning space (Esper 1966; Kirby et al. 2008) establishing isomorphism between semantics and syntax—i.e., a systematic mapping between the parts of the meaning and those of the form (Montague 1970). These results replicate the results found in previous IALL studies (Kirby et al. 2008, 2015; Silvey et al. 2015) but with a more complex and substantial meaning space.

Our results further suggest the evolution of morphologically complex constituents which constitute a nominal syntactic category; they all share the same distribution within sentences and thus can be interchanged with each other to derive grammatical structures. Moreover, all nominal constituents within a given language share a morphological category marker and thus high string-similarity (for similar results, see Carr et al. 2016; Nowak & Baggio 2016). These nominal constituents combine with each other and verbal elements to form even more complex linguistic expressions. Compared to previous IALL studies, this is the most productive structure hitherto shown to evolve. Here I show at least two levels of the hierarchy of constituent types in natural languages discussed in 2.1.1: morphs combine to form word-like forms and these further combine to form sentence-like structures. It is thus the first time an IALL study shows the evolution of more than one level of constituency (i.e., lexical items—which also constitute phrases, and sentences) and can describe nominal elements as nodes within hierarchical constituent structure and not just isolated referring expressions or labels which can be obtained via basic concatenation. Additionally, I show the evolution of word order rules in argument structure to encode the semantic roles in motion events. As languages become more structured, the order of nominal constituents in a sentence also becomes more systematic: the position in which the same nominal constituent appear determines the argument they refer to. A recent study by Nowak and Baggio (2016) showed the emergence of word order regularity in a multigenerational signalling game, but given that their objects can only appear either in subject or object position but not both, “positional” compositionality cannot be evidenced. This study further supports the emergence of word order regularity through cultural evolution in the

laboratory and provides evidence for positional compositionality.

Nevertheless, linguistic structure only fully evolved in half of the languages: whilst nominal constituents encoding the shape and the number of the objects evolved in all languages, verbal constituents that systematically mapped to the motion and aspect features of scenes (which occur less frequently) only evolved in two. Correspondingly, although complex nominal syntactic categories evolved and word order regularity increased across languages, it was only in half of the languages that these properties were consistently systematic throughout the last generations of a diffusion chain.

It is possible that the restrictiveness of the pressure for expressivity might hinder the evolution of structure in certain cases. I introduced a pressure for expressivity into the experimental model via a highly restrictive filter against ambiguity, which prevented languages to become degenerate (i.e., it guarantees that each sentence corresponded to a single scene). Every time a participant used the same description for more than one meaning thus introducing homonymous descriptions in the system, they were warned and asked to provide an alternative description. This and similar artificial filters against ambiguity (Carr et al. 2016; Kirby et al. 2008; Silvey et al. 2015; Verhoef 2012) have been previously used as an analogue of a pressure to be expressive which comes from the need to communicate accurately in natural language use. Nonetheless, with a complex meaning space where the discriminating features of meanings might not be clear to the participant, an artificial pressure such as the one implemented might cause participants to add linguistic forms which do not map to any specific meaning feature. Figure 2.9 shows that participants in chains A2 and A4—where we observed most unsystematicity—often struggle to discriminate meanings and produce homonymous sentences before they are asked to introduce an alternative description. Without the need to communicate meanings accurately, participants who do not discriminate all meaning features systematically do not have any natural reason to do so in production, and the artificial pressure forces participants to add or delete elements in linguistic expressions which do not necessarily map to any semantics.

In order to be successful, communication (i.e., the natural promoter of expressivity) requires a shared communication system between interlocutors (Lewis 1968; Schelling 1960). Every communicative interaction requires speakers to convey a meaning producing an utterance and hearers to arrive at the correct interpretation of it. In order to meet this requirement interlocutors have to coordinate to establish a set of shared conventions. Thus on top of a pres-

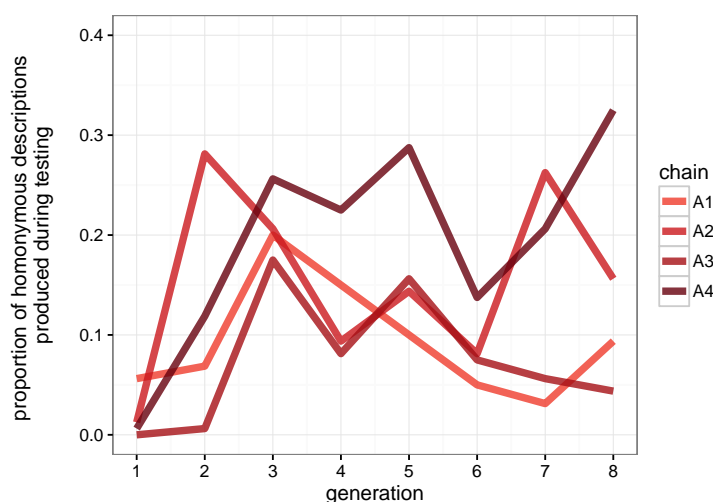


Figure 2.9: Proportions of homonymous descriptions introduced by participants during testing in Experiment 1, and thus the proportions of trials (out of 160) in which participants were asked to provide an alternative description as they had used it previously to refer to a different scene. We observe that whilst in chains A1 and A3 the proportion of homonyms introduced decreases in the last generations, the contrary tendency is found in chains A2 and A4.

sure for expressivity, communication introduces a pressure for coordination (Winters 2017): interlocutors need to align on a shared system which is both expressive and learnable, otherwise they will not be able to discriminate between meanings or be able to express new ones. A pressure for coordination further increases the benefit of compositional structure as it allows interlocutors to understand each other productively in the simplest way, i.e., by predictable associations between syntax and semantics (Pagin 2012). It is then possible that the presence of a coordination pressure together with expressivity and learnability pressures facilitates the evolution of compositional structure.

Interaction studies have shown that iterative communicative interaction, with feedback provided (Krauss & Weinheimer 1966), leads to the establishment of conventions and the successive simplification and systematisation of communicative systems (H. H. Clark & Wilkes-Gibbs 1986; Garrod, Fay, Lee, Oberlander, & MacLeod 2007; Selten & Warglien 2007) (for review, see Galantucci & Garrod 2011). More recent work has also framed alignment in communicative interaction as an alternative source of regularisation of inconsistencies in language (Fehér, Ritt, & Smith 2017; Fehér, Wonnacott, & Smith 2016). These results hint that there is also a learnability pressure acting during coordination, i.e., in the construction of a shared system of conventions between interlocutors, as they learn about each other's systems. Experimental work integrating both transmission with communicative interaction to explore the

evolution of graphical systems has shown that some degree of structure does emerge through interaction alone (Theisen et al. 2010). However, it does not accumulate over time as it does with the inclusion of transmission and thus the systems that evolve from iterative interaction alone are not as systematic or structured (Theisen et al. 2010; Theisen-White et al. 2011). Moreover, within the linguistics modality, Carr et al. (2016) found that systematic (although not compositional) structure can evolve from both transmission and interaction together but not from transmission and a filtering against ambiguity.

Altogether, these studies suggest a non-trivial effect of communicative coordination in the evolution of language, be it for the establishment of shared conventions or for its role in increasing systematicity when supplemented by a transmission bottleneck. In Experiment 2, I explore whether the substitution of an artificial pressure for communicative interaction facilitates the evolution of complex compositional structure.

2.3 Experiment 2: transmission and communication

In Experiment 2 I utilise the methodology used in Kirby et al. (2015) and Winters et al. (2015) and introduce a more naturalistic pressure for expressivity through the implementation of communicative interaction at each generation in transmission chains. I run four transmission chains of 8 generations each. In overview, pairs of participants at each generation try to learn an artificial language based on some linguistic data, and then during testing they use it to communicate with each other, producing a set of linguistic data which will be the input for the next pair of participants in the chain.

2.3.1 Method

2.3.1.1 Participants

Sixty-four participants were recruited to participate in an communication game through the University of Edinburgh Careers Service database of student and graduate employment. All participants were native speakers of English (mean age 22 years, age range 18–35); each received a payment of £9. The experiment was conducted in accordance with the ethics procedures of Linguistics and English Language, The University of Edinburgh.

2.3.1.2 Procedure

Training and communication In Experiment 2, pairs of participants were asked to individually learn an artificial language which later they would use to communicate with each other. I used the same stimuli as in Experiment 1 (see 2.2.1.2) and initial languages were generated as for Experiment 1 as well (see 2.2.1.3)¹².

Experiment 2 was divided into two phases: a training phase and a communication phase. During training, two participants were trained in parallel on a set of 40 out of the 80 pairings contained in the full language¹³. The two participants in each dyad were trained on the same set of 40 pairings, balanced to contain at least one instance of all meaning features and feature values. They saw each item in the training set three times (order randomised for each participant), giving a total of 120 training trials. After each training trial, participants underwent the same type of recall tests described in section 2.2.1.4 for Experiment 1: participants were asked to type in the descriptions they were just presented with (50% of the time) or they were asked to select the scene they just saw in the trial (other 50% of the time).

During communication, the pairs of participants were asked to communicate with each other using the language they had just learned. There were two roles participants played in this stage, sender and receiver. Pairs communicated the whole set of 80 stimuli during the testing phase, each participant communicating a subset of 40 (again balanced to contain instances of all the different possible values of each feature). Participants swapped roles at every trial. The sender was presented with a scene for 5 s (without a description) and then was asked to type in a description for that scene. The description was then sent to their partner, the receiver. The receiver had to identify the scene the sender described by selecting a scene out of four different ones displayed in a two by two grid (the target scene and three randomly chosen foils from the set of 80 pairs). Full feedback was provided after each trial: participants saw a screen with a red or green background—depending on the communicative success (green for success and red for failure)—which displayed the description the sender typed in alongside the meaning

¹²I nevertheless added a few restrictions to the initial languages generated in Experiment 2 that were not present in Experiment 1: I excluded the character <s> in order to avoid its use as a plural marker (as seen in a language A3 in Experiment 1) and <k,q,w,x,y,z> were further excluded in order to avoid that participants notice the restriction. Characters absent in the initial languages were blocked in the keyboard and participants could not enter them in their responses.

¹³Here they were only trained in 40 items because all the data came from only one participant in the previous dyad. Each participant produced 40 items in the test phase as having each producing 80 items would have taken too long (approx. 2 hours).

the sender was trying to convey and the meaning the receiver selected. Feedback is one of the main resources from which alignment between interlocutors draws upon (Garrod & Pickering 2009; Spike et al. 2017). By providing participants with full feedback at each communicative event I allow participants to test their predictions and refine them to match their interlocutors' scene-description mappings in future trials.

Transmission Pairs of participants were organised into independent transmission chains and the transmission procedure was implemented as per Experiment 1. At each generation, I randomly selected one participant's set of productions out of the pair (composed of 40 scene-description pairings) and used it as the training language for the next generation.

2.3.2 Analyses and results

2.3.2.1 Languages

As for languages in Experiment 1, I extracted a matrix of associations between lexical items and their referents in the scenes as explained in section 2.2.2.2. Figure 2.10 shows the different word-value(s) associations that form the lexicon of the example language C2 (where the C stands for Communication plus transmission and 2 indicates the chain number). Examples showing the arrangement of morphs within descriptions are provided in (6).

- (6) a. roji ref tube evoto ref
square group slide circle.group group
'A group of squares slid towards a group of circles'
- b. evo-to ref tube tube roji
circle.group group slide.ongoing square
'A group of circles slide towards a square back and forth'
- c. roji babatube babatube evo
square bounce.ongoing circle
'A square bounces towards a circle back and forth'

In Tables 2.4 and 2.5 I show the different nominal and verbal morphology from all final languages in Experiment 2. As in Experiment 1, Shape and Number are encoded within a

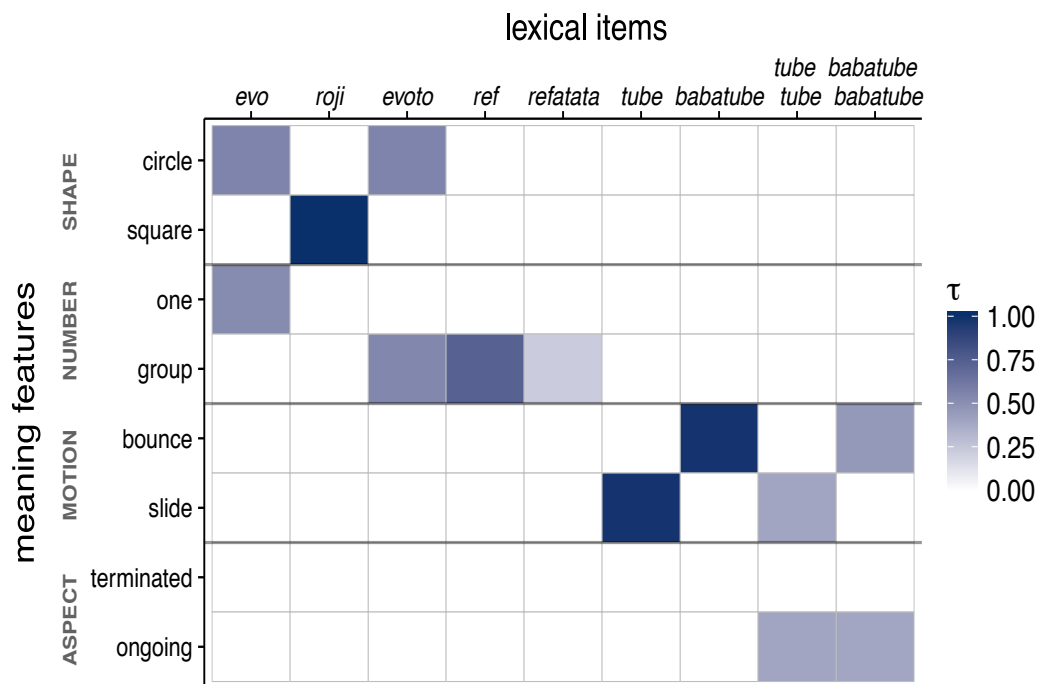


Figure 2.10: Lexical items in language C2. From left to right, we find that there are two nominal lexical items associated with Shape: *roji* ('square') and *evo* ('circle'). Plurality is generally marked by the morph *refatata* or its clipped pair *ref* (unconditioned variation) after nominals to form a complex nominal (e.g. *roji ref* or *roji refatata* ('a group of squares')). Nevertheless, *evo* takes also a bound morph *-to* and forms plural with *evoto ref* ('a group of circles') (see the examples of sentences in (6) in the main text). We also observe two verbal elements associated with the feature of Motion: *babatube* ('bounce') and *tube* ('slide'). Their default Aspect is *terminated* if they appear on their own, and the 'ongoing' aspect is marked by full reduplication of the verbal elements: *babatube babatube* ('ongoing bouncing') and *tube tube* ('ongoing sliding').

		Lang C1	Lang C2	Lang C3	Lang C4
circle	one	po	evo	to/ce-	domo
	group	popo	evo-to ref	cecede/ceci-	domoge
square	one	vahu	roji	me/me-	pira
	group	vahuvahu	roji ref	mecede/meci-	pirage

Table 2.4: Nominals within final languages in Experiment 2. Nominals in Language C3 can be expressed through free morphs (between white spaces) as well as bound roots, separated by a slash in this table. Roots take on suffixes marking aspect. Only free morphs can appear as anchor arguments, whereas both free morphs and bound roots can appear in focal arguments.

		Lang C1	Lang C2	Lang C3	Lang C4
slide	ongo.	jiji-F A	F tube A	F-jijiju mu A	F refugo A
	term.	ji-F A	F tube tube A	F-jiju mu A	F lefugo A
bounce	ongo.	jiji-F A	F babatube A	F-jijiju ju A	F refugo A
	term.	ji-F A	F babatube babatube A	F-jiju ju A	F lefugo A

Table 2.5: Verbal elements of the final languages in Experiment 2. Focal (F) and Anchor (A) indicate the most frequent position of the nominal arguments in a description.

complex nominal (see Table 2.4). Plurality is expressed via full reduplication (C1), a free morph (C2), and suffixation (C2, C3 and C4). All of the languages except for C4 mark *ongoing* Aspect of an event via full reduplication. In C1 and C3 it is the marker for a *terminated* event that is reduplicated, whereas in C2 we observe the full reduplication of the forms encoding Motion. Word order is fixed across languages: focal arguments precede anchor arguments consistently (see section 2.3.2.4).

Unexpectedly, we observe that half of the languages are underspecified: i.e., languages C1 and C4 are underspecified for Motion —i.e., they do not distinguish between *bounce* and *slide*. Figure 2.11 shows the number of distinct descriptions in language as well as their number of occurrences, and thus the number of different motion event scenes they refer to. We observe that in languages C1 and C4 most of the descriptions refer to two different meanings (i.e., they are homonymous). Although underspecified, the languages are highly systematic and participants' communicative accuracy scores are high: for the languages shown in Tables 2.4 and 2.5, dyads communicate successfully with $\hat{p} > 0.95$, see Figure 2.12. Participants thus communicate successfully without the specification of each meaning feature; since the foils in discrimination arrays were selected randomly in all matching trials, in most cases specifying the focal and anchor objects was sufficient to disambiguate.

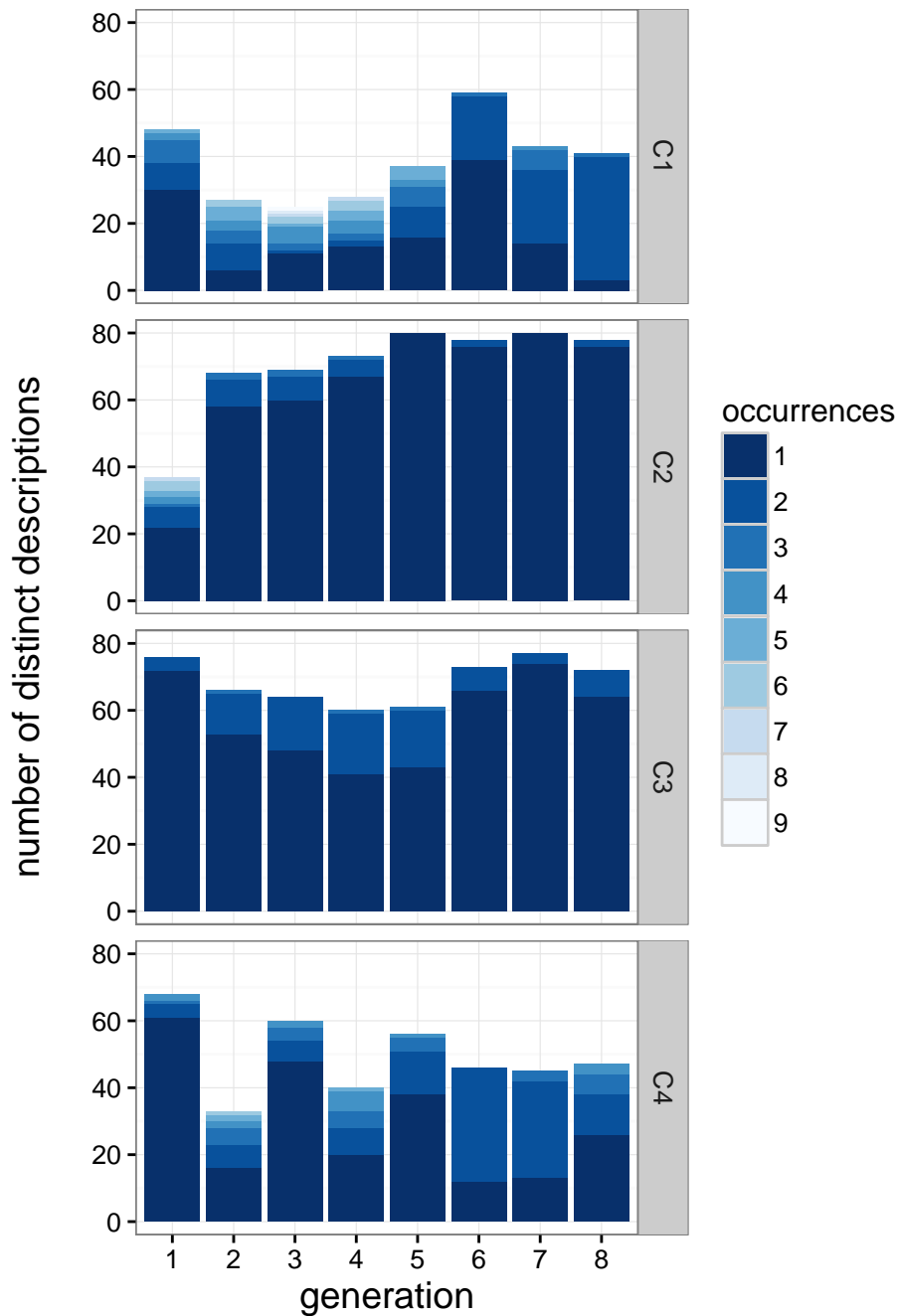


Figure 2.11: Number of distinct descriptions and their occurrences in a language. A fully expressive language would have 80 distinct descriptions, one per scene. Any description that occurs more than once in a language introduces ambiguity into the system. Languages C2 and C3 at the final generations are fairly expressive, most of their descriptions only occur once and thus are only associated with one scene. By contrast, the final languages C1 and C4 are underspecified and thus less expressive: most of the descriptions are homonyms often corresponding to two different scenes (i.e., corresponding to the observed underspecification of either Motion or Aspect meaning features).

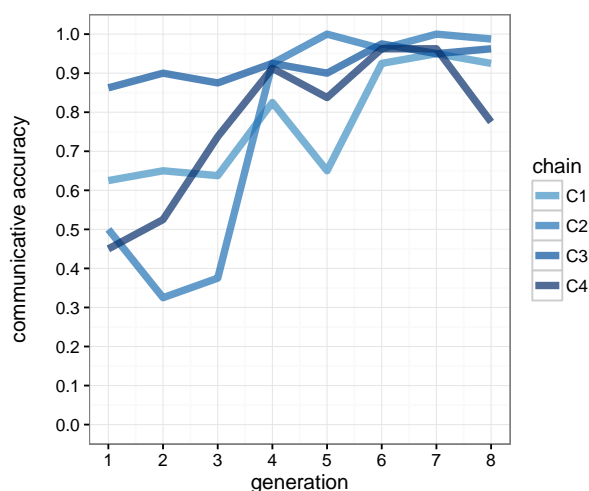


Figure 2.12: This graph shows the communicative accuracy as a proportion of the successes during communication (80 trials) between pairs of participants in Experiment 2. We observe an increase in communicative accuracy in the first five generations, where it is on average $\hat{p} > 0.8$. It later continues on increasing and it is $\hat{p} > 0.9$ in all languages by the final generations. Exceptionally, communicative accuracy drops to $\hat{p} = 0.775$ in the last generation of chain C4, where we also observed a drop in linguistic structure.

2.3.2.2 Compositional structure

Figure 2.13a shows the structure scores obtained in the experimental data. As in Experiment 1, I performed a segmented linear mixed-effects model with a breakpoint in generation 4 (obtained as per Experiment 1) to explore the effect of generation on linguistic structure across experiments. We will call this Model 4. As fixed effects I entered Generation with Indicator nested as well as Experiment (Experiment 1 and Experiment 2). This equates to the introduction of Generation, Experiment, the interaction between Generation and Indicator, and the interaction between Generation, Indicator and Experiment. For all models reported, I use reverse Helmert contrasts for the fixed effect Experiment; the intercept is the mean of Experiment 1 and Experiment 2, and Experiment 2 is compared to Experiment 1. As random effects, I introduced intercepts for Chain as well as by-Chain slopes for the effect of Generation. The overall model fit was $R^2_{\text{marginal}} = 0.618$ and $R^2_{\text{conditional}} = 0.676$. Figure 2.13b shows the fixed and random estimates obtained in Model 4. The model intercept indicates that languages were highly structured by generation 4 ($\beta = 19.689, SE = 1.931, p < 0.001$). A significant effect of Experiment ($\beta = 6.850, SE = 1.931, p = 0.014$) suggests that languages in Experiment 2 were significantly more structured at generation 4 than languages in Experiment 1. I found a significant effect of Generation ($\beta = 4.763, SE = 0.727, p < 0.001$) and a marginally significant effect of the in-

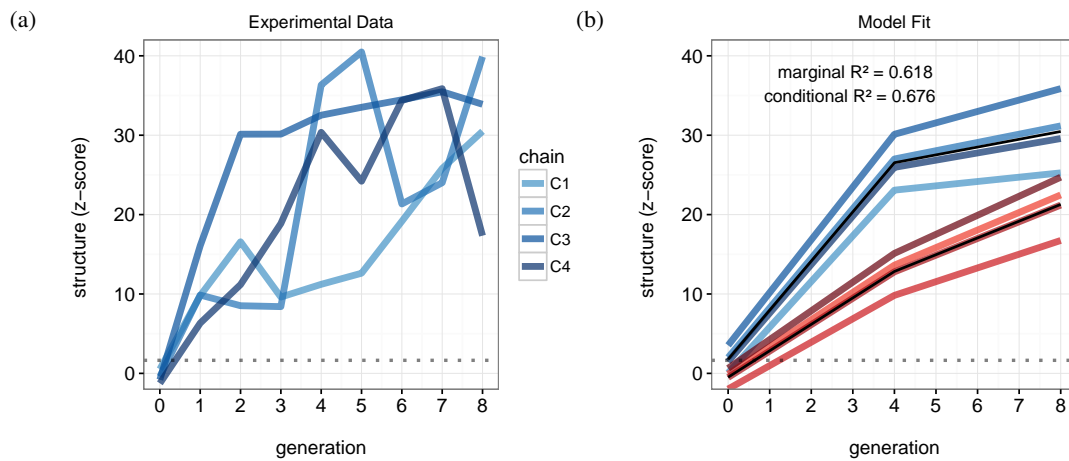


Figure 2.13: (a) Linguistic structure over generations for each of the four transmission chains in Experiment 2. Linguistic structure increases as languages are transmitted through generations of learners. (b) Fitted values from the mixed-effects regression Model 4 for the four transmission chains in Experiment 1 (red) and the four transmission chains in Experiment 2 (blue). Coloured lines represent the random slopes estimates (for generation) depending on random intercepts (individual chains), whereas the black lines represent the fixed effects estimates for each experiment. In both plots, the dotted horizontal line represents the chance level (z -score 95%CI = ± 1.645 , one-tailed).

teraction between Generation and Experiment ($\beta = 1.431, SE = 0.727, p = 0.053$), suggesting that although structure increased in the first four generations across experiments, the increase was significantly greater in Experiment 2. I also found a significant interaction between Generation and Indicator ($\beta = -3.216, SE = 1.275, p = 0.014$) and no effect of the interaction between Generation, Indicator and Experiment ($\beta = -1.994, SE = 1.275, p = 0.123$), suggesting that structure increased less by generation in the second half of transmission chains in both experiments¹⁴. Altogether, these results suggest that languages in Experiment 2 become more structured faster, but that structure increases by generation across both experiments. However, this effect of generation on linguistic structure is not equally pronounced across generations in transmission chains; slopes were steeper initially and less so further in the chain as languages became more stable as a result of the cumulative increase in structure. These non-linear evolutionary trajectories are reminiscent of the (more or less pronounced) logarithmic curves shown in many models of cultural evolution (Boyd & Richerson 1988; Claidière & Sperber 2007; Griffiths, Kalish, & Lewandowsky 2008; Henrich & Boyd 2002; Mesoudi 2011).

¹⁴Note that although Model 1 did not suggest a steeper slope for structure in the first half of chains in Experiment 1, Model 4 suggests that, in both experiments, slopes are steeper in the first half of chains.

2.3.2.3 Hierarchical constituent structure: the emergence of complex nominal constituents

Table 2.6 shows the relative frequency with which morphs encoding Shape and Number appear next to each other in a language. In all four chains, we consistently find complex nominal constituents already by generation 1. As described in section 2.3.2.1, either Number was marked via reduplication or morphs encoding Number followed morphs encoding Shape. As in Experiment 1, I tested whether these complex nominals in fact constitute a syntactic category. Figure 2.14a shows the distributional distance of nominal constituents; scores below -1.645 indicate that the distributional similarity between nominal elements is unlikely to result by chance. All final languages in Experiment 2 obtain z-scores below chance level and thus I conclude that nominal syntactic categories evolve via cultural transmission. I also performed a linear mixed effects model, which I will call Model 5, to test the effect of generation on the distributional distance in across experiments. As fixed effects, I entered Generation (centred) and Experiment (Experiment 2 vs. Experiment 1) as well as their interaction. As random effects, I introduced an intercept for Chain and a by-Chain random slope for the effect of Generation. The overall model fit was $R^2_{\text{marginal}} = 0.459$ and $R^2_{\text{conditional}} = 0.553$. Figure 2.14b shows the fitted values of Model 5 for fixed and random effects. Results showed a significant effect of Generation ($\beta = -0.327, SE = 0.061, p < 0.001$) and no significant interaction between Generation and Experiment ($\beta = 0.040, SE = 0.061, p = 0.5371$), suggesting that the distributions in which nominals appear became more similar by generation to a similar degree across experiments. I did not find an effect of Experiment either ($\beta = -0.027, SE = 0.115, p = 0.819$), suggesting that both experiments obtained similar estimates at the intercept (between generation 4 and 5). It is worth noting that larger z-scores can be obtained from languages with larger lexicons because the probability of selecting morphs that are nominals during repeated random sampling is lower (see section 2.2.2.3). Because languages in Experiment 1 have larger lexicons (primarily because they are not as structured and systematic as languages in Experiment 2), z-scores obtained in the two languages that evolve nominal syntactic categories are lower (see section 2.2.3.3), leading to an average z-score similar to that in Experiment 2. There is a clear difference between experiments in the amount of languages in which we observe the evolution of a nominal syntactic categories; this only in two in Experiment 1 and in all four in Experiment 2 instead. Nevertheless, we observe that distance between nominals diminishes by generation

	gen 1	gen 2	gen 3	gen 4	gen 5	gen 6	gen 7	gen 8
chain C1	1	1	1	1	1	1	1	1
chain C2	1	1	0.99	1	1	1	0.96	0.99
chain C3	1	1	1	1	1	1	1	1
chain C4	1	1	1	1	1	1	1	1

Table 2.6: Proportion of adjacent Shape and Number morphology for a specific object, either focal or anchor. Shape and Number are always encoded adjacent to each other in all languages from the first generation onwards.

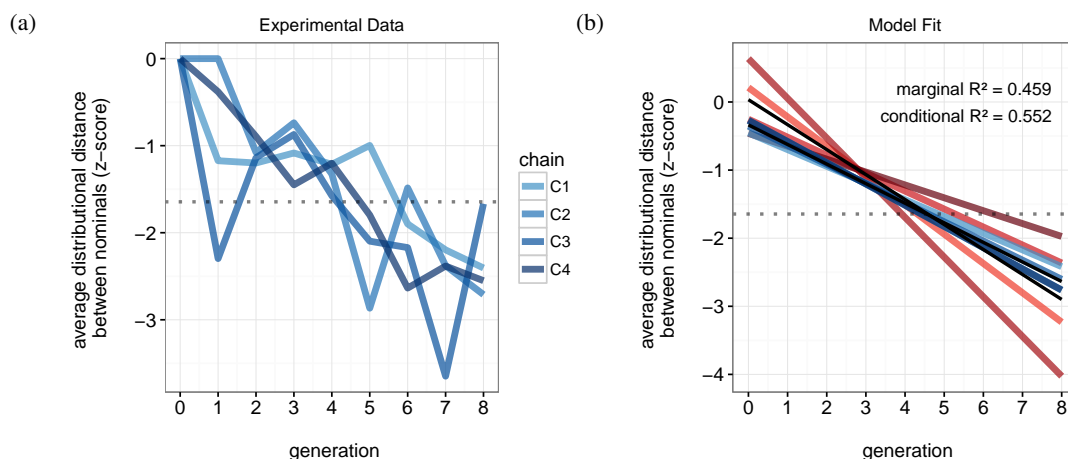


Figure 2.14: (a) Distributional distance between a language’s nominals through generations for each of the four chains in Experiment 2. The dotted line represents the chance level (z -score 95%CI = ± 1.645 , one-tailed); z -scores below it indicate that the distributional similarity between nominals is unlikely to arise by chance. Distributional distance between nominals decreases with generation as languages become more structured, suggesting the emergence of a nominal syntactic category. We observe that all the average distance between nominals are below chance level in the last two generations. (b) Fitted values from the mixed-effects regression Model 5 for the four transmission chains in Experiment 1 (red) and the four transmission chains in Experiment 2 (blue). Coloured lines represent the random slopes estimates (for generation) depending on random intercepts (individual chains), whereas the black lines represent the fixed effects estimates for each experiment.

across all languages in both experiments.

2.3.2.4 Word order rules for nominal arguments

The stacked area graphs in Figure 2.15 show the proportions of focal-anchor, anchor-focal and undefined orders at each generation for each of the four transmission chains in Experiment 2. As explained in section 2.2.2.4, these proportions are taken from the set of 48 descriptions in which two different objects or sets of objects occupy focal and anchor semantic roles. We observe that the proportion of undefined order decreases as languages are transmitted through generations, and the proportion of focal-anchor orders increases rapidly. Moreover, we observe

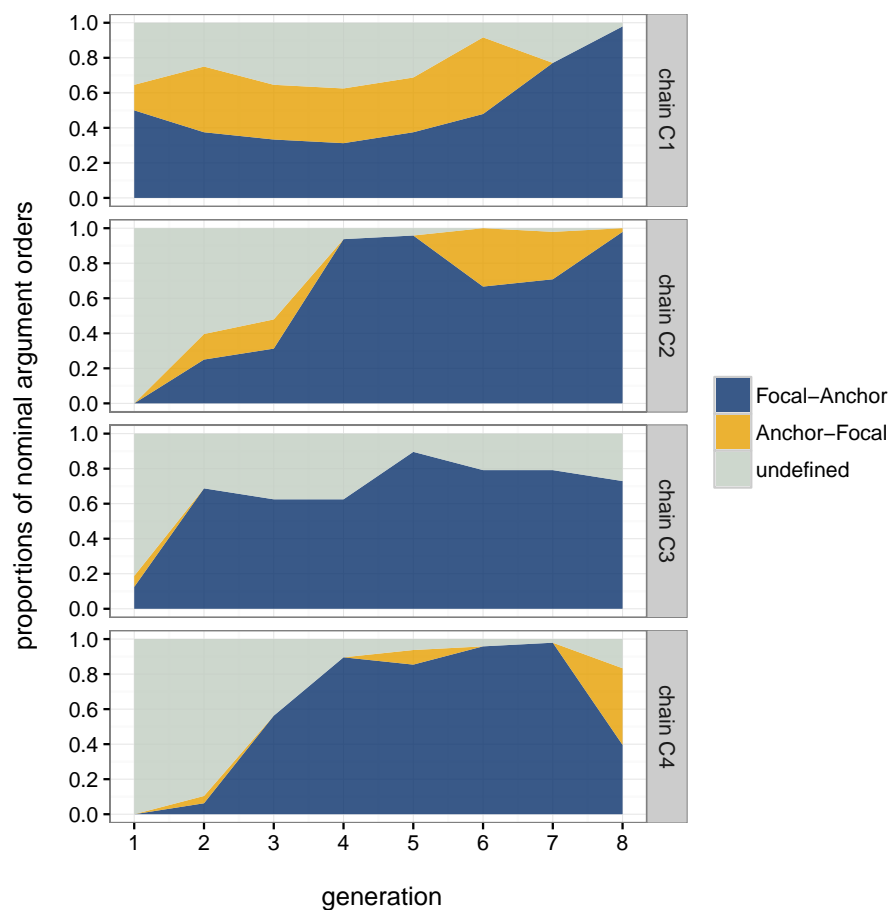


Figure 2.15: Proportion of word order types for nominal arguments by chain and generation. A fixed Anchor-Focal word order rule evolves across chains.

that the order of nominal arguments is mostly fixed in the last generations. Word order regularity shows yet another aspect in which languages evolved in Experiment 2 are more systematic.

I ran a linear mixed effects model, which I will call Model 6, to test the effect of generation on the variability of nominal argument orders in languages now in Experiment 2 as well as in Experiment 1 —calculated by the entropy of the system of orders as described in section 2.2.2.4. I entered Generation (centred) and Experiment (Experiment 2 vs. Experiment 1) as well as an interaction term between them. As random effects I entered intercepts for Chain as well as by-Chain slopes for the effect of Generation. Figure 2.16 shows the nominal order variability scores of the experimental data (Figure 2.16a) as well as the fitted values of Model 2 for fixed and random effects (Figure 2.16b). The overall model fit was $R^2_{\text{marginal}} = 0.362$ and $R^2_{\text{conditional}} = 0.384$. Results show a significant effect of Generation ($\beta = -0.073, SE = 0.014, p < 0.001$) and no significant interaction between Generation and

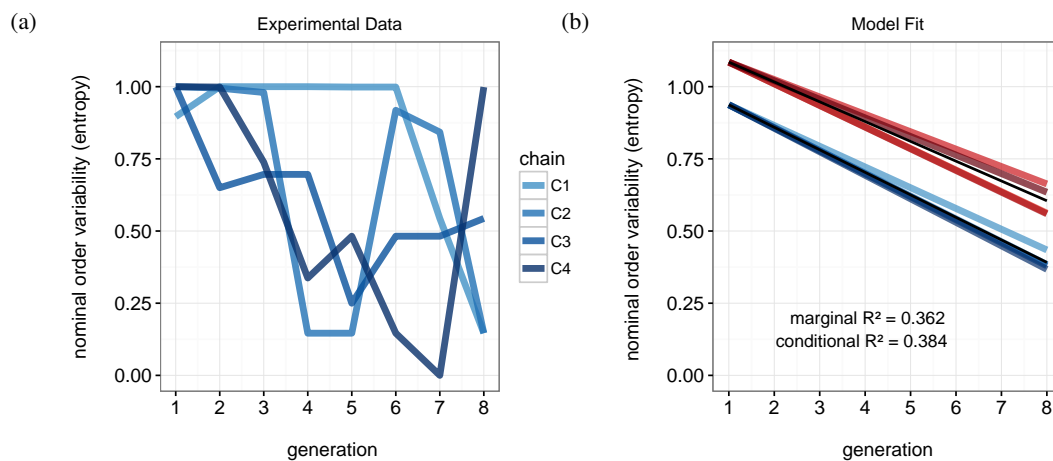


Figure 2.16: (a) Variability of nominal argument orders by generation and chain in Experiment 2. (b) Fitted values from the mixed-effects regression Model 6 for the four transmission chains in Experiment 1 (red) and the four transmission chains in Experiment 2 (blue). Coloured lines represent the random slopes estimates (for generation) depending on random intercepts (individual chains), whereas the black lines represent the fixed effects estimates for each experiment.

Experiment ($\beta = -0.005, SE = 0.014, p = 0.744$), suggesting that entropy decreases by generation to a similar degree across experiments, and therefore that the order of nominal arguments becomes more consistent as languages are transmitted through generations of learners. I also found a significant effect of Experiment ($\beta = -0.090, SE = 0.034, p = 0.022$) indicating that the order of nominals is less consistent in the languages in Experiment 1 at the intercept (between generations 4 and 5).

2.3.3 Discussion

2.3.3.1 The evolution of complex compositional structure

So far in this study I have manipulated the nature of a pressure for expressivity (artificial vs. communication) whilst keeping constant a learning bottleneck in transmission, which promotes the need for generalisation and thus for the learnability of the language. I have shown that compositional structure evolved across experiments through cultural transmission thus replicating the results found in previous IALL studies (Kirby et al. 2008, 2015). However, only some aspects of compositional structure had been shown hitherto (Kirby et al. 2008, 2015; Nowak & Baggio 2016; Silvey et al. 2015; Winters et al. 2015), namely what I have referred to as basic compositionality: simple forms map to simple meanings and these combine by means

of concatenation (left or right), without further syntactic or semantic structure. The type of compositionality is thus confined to isolated lexical items. With the present study we can now add two more aspects of compositional structure which can emerge through IALL, namely its hierarchical constituent structure and its “positional” aspect—given by word order rules which determine the semantics of constituents. These two aspects together highlight the relevance of the second part of the compositionality axiom (i.e., “the meaning of the whole is determined by the meaning of its parts and *the way they are combined*”).

As in Experiment 1, the languages that evolved structure thus comply with the characteristics of configurational languages where word order is fairly fixed and sentences are mainly composed of morphosyntactically continuous expressions (i.e., continuous constituents, without long-distance dependencies). Moreover, more frequent and salient meaning features such as Shape (Gentner 1982; Landau, Smith, & Jones 1988) and its associated feature Number were always encoded, unlike Motion and Aspect which were not always both encoded within a language. And although the strategy for marking number varied across the evolved languages in this study (e.g., simulfixation, suffixation, reduplication, suppletion or plural word), morphs encoding Shape and Number appeared always adjacently to each other, forming continuous nominal constituents. This is consistent with the universal tendency (and with English grammar) to mark nominal number in the Noun or in its immediate periphery (Dryer 2013a). Nominal constituents not only evolved across all languages without exception, they also emerged early on in the chains. This conforms to a noun bias parallel to that suggested in language acquisition in many languages (Dhillon 2010), which we can define for early language formation as follows: a systematic nominal category emerges earlier than other categories (for a similar result in the gestural modality, see Motamedi, Schouwstra, Smith, Culbertson, & Kirby under revision). Moreover, it is also worth noting that where fixed word order evolved, regardless of the position of motion and aspect lexical items within the system (i.e., whether it appeared sentence medial or final), focal objects always precede anchor objects, consistent with the universal Agent-first tendency in natural languages (including English, the native language of the participants) (Dryer 2013b; Greenberg 1966) and a universal processing bias (Gibson 2000; Hawkins 2004; Marantz 2005).

However, results also show that the nature of the pressure for expressivity determines the evolutionary rate of linguistic structure. Whilst all languages that evolved in Experiment 2

evidenced complex compositional structure and were (almost) perfectly systematic, this was not true of all languages in Experiment 1.

2.3.3.2 The effect of communicative interaction in the evolution of complex linguistic structure

Although similar linguistic structure evolved across experiments, it did so to varying degrees depending on the nature of a pressure for expressivity, i.e., either an artificial pressure against ambiguity in production or communicative interaction. Results show that the substitution of an artificial pressure for communicative interaction eases the evolution of linguistic structure. With the inclusion of communicative interaction, languages become significantly structured by the first generation already, which signals a more rapid emergence of structure than in Experiment 1. Moreover, all languages in Experiment 2 evolve to be significantly more structured and systematic: descriptions only contain morphology which has a semantic mapping to the constituent parts of the meaning they describe and word order is fixed. The only aspect in which languages that evolved in Experiment 1 are more systematic than those in Experiment 2 is in the sublexical structure within morphology encoding Shape: languages in Experiment 2 do not evolve nominal category markers and thus, string-similarity between nominals is lower within a language.

These differences observed between conditions suggest that the expressivity that communication promotes is not analogous to an artificial pressure against ambiguity in a language—at least given the presence of a complex meaning space and the highly restrictive nature of the artificial pressure. As a matter of fact, half of the languages that evolved in Experiment 2 were underspecified for one meaning feature (either Motion or Aspect, but never Shape or Number). This suggests that provided that not all meaning features are required to be discriminated at every communicative event, communicative interaction does not impose as strong a pressure for expressivity as assumed in Experiment 1. In Experiment 2 participants are often not required to discriminate all features of the meanings for communication to be successful. In natural languages it is not necessary (or logically possible) to specify all aspects of a meaning in a concrete communicative event—be it because they are provided by the context or they are simply not required; therefore, it is more economical or at least sufficient to encode the minimum meaning features, minimising the effort of unambiguously conveying a message (Brochhagen,

Franke, & van Rooij 2016; Winters et al. 2015). The differences between the underspecification found in Experiment 2 compared to the full expressivity found with a similar design in Kirby et al. (2015) is most probably due to the differing complexity of the meaning space and the size of the context array the receiver has to select meanings from during communication. Whist in Kirby et al. (2015) participants are asked to select an object out of a context array of 6 with a much simpler meanings space (i.e., 12 objects in total, only differing in shape and fill-pattern), participants in Experiment 2 are only asked to discriminate the scene conveyed by the partner out of an array of four scenes (randomly selected) at each communication trial and with a substantially more complex meaning space (i.e., 80 meanings, with 5–7 features each one). The probability of having to discriminate every single meaning feature value of a scene is lower in my design than it is in Kirby et al. (2015).

Altogether, these results suggest that a coordination pressure in communicative interaction contributes significantly to the emergence of linguistic structure. The possibility of coordination between participants is ultimately what speeds it up and leads to a very early emergence of structure in Experiment 2 (see also Winters 2017). With a shared goal to communicate accurately, participants prioritise to establish communicative pacts with partners to bootstrap communication—even at the expense of faithful reproduction of the learned language. It is probable that the inclusion of communication and thus of the explicit goal of arriving at a shared system for communication results in conscious design by language users more than in Experiment 1. Nevertheless, the degree of structure increases as languages are transmitted through a learning bottleneck and thus similarly to Theisen et al. (2010) and Theisen-White et al. (2011) in the graphic modality, results show that a certain degree of structure can emerge during communicative interaction but only through iterated learning does it accumulate.

In sum, the addition of communicative interaction to transmission facilitates the evolution of linguistic structure. The effect of communication in this study is thus not comparable to an artificial pressure against ambiguity in production. The coordination pressure at play during communication facilitates the conventionalisation of lexical items and grammatical rules. Moreover, communicative interaction (i.e., without a requirement of full discrimination at each communicative event) does not impose such a hard constraint on expressivity as the one assumed in the artificial pressure against ambiguity: most languages are underspecified for one meaning feature. This can be explain given that underspecification minimises effort in produc-

tion, and communicative effectiveness was not compromised. In Experiment 3, we explore the effect of communicative interaction supplemented with an artificial pressure against ambiguity.

2.4 Experiment 3: transmission, communication and artificial pressure against ambiguity

Given the underspecification and the simplicity of the languages obtained in Experiment 2, I designed a further Experiment 3 which incorporates both communication and an artificial pressure against ambiguity to help participants repair underspecification and increase the complexity of linguistic structure. I expect that the added artificial pressure leads to the evolution of more expressive and complex languages than in Experiment 2, and that the addition of communication leads to earlier and higher levels of structure than in Experiment 1. If this is the case, I expect that the combination of pressures alters the evolutionary rate of linguistic structure in comparison to the other two experiments. Given that more conventions would need to be established between interlocutors, I expected structure to evolve at a slower rate in Experiment 3 than in Experiment 2, but earlier than in Experiment 1 (which does not include communicative interaction). However, we should find that structure scores are higher in Experiment 3 than in Experiment 2 eventually, since the artificial pressure will prevent underspecification in languages—which will necessarily require more morphs and/or grammatical rules and thus more structure.

2.4.1 Method

2.4.1.1 Participants

Sixty-four participants were recruited as per Experiments 1 and 2. All participants were native speakers of English (mean age 21.22 years, age range 18–30). Participants received a payment of £9. The experiment was conducted in accordance with the ethics procedures of Linguistics and English Language, School of Philosophy, Psychology and Language Sciences, The University of Edinburgh.

2.4.1.2 Procedure

Experiment 3 followed the same procedure as Experiment 2 (transmission and communication, see section 2.3.1.2) with the addition of the artificial pressure against ambiguity in production used in Experiment 3 (transmission and artificial, see section 2.2.1.4). When acting as a sender during communicative interaction, participants were not allowed to send descriptions to their partner that the pair had already used for another meaning. As in Experiment 1, if senders typed in descriptions already used, they were told by the computer that the same description was already in use for another meaning and they were asked to type in a different description.

2.4.2 Analyses and results

2.4.2.1 Languages

Figure 2.17 presents the association matrix between lexical items and meanings of language CA1 (where CA stands for Communication plus Artificial pressure and 1 indicates the chain number). In the example sentences provided in (7) below we can see the syntactic organisation of these lexical items.

- (7) a. *deeja reeva bo deju*
 square.group slide circle.group
 ‘A group of squares slid towards a group of circles’
- b. *deju reeva deeja*
 circle bounce squares
 ‘A group of squares bounced towards a circle’
- c. *deeja reeva deju*
 square.group slide circle
 ‘A circle slides towards a square back and forth’

Table 2.7 shows the nominal elements of the four languages that evolved in Experiment 3. We show the language of the final generations for all but CA2, for which we show the language of the penultimate generation before it abruptly decreases its systematicity at the final generation (see the following sections 2.4.2.2–2.4.2.4). Note that full or partial category marking within nominal forms to distinguish Shape is a common trait amongst languages in

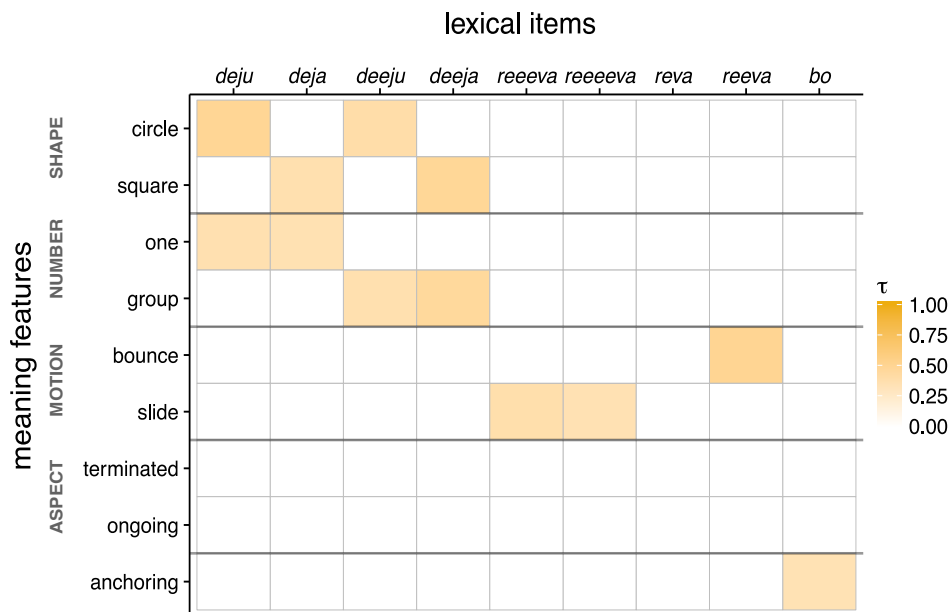


Figure 2.17: Lexical items in the final language CA1 (chain 1) in Experiment 3. From left to right, we observe nominal elements associated with Shape and Number, verbal elements associated with Motion and an apparent functional element associated with Anchoring. The two nominal forms *deju* and *deja* stand for *circle* and *square* respectively. Singularity is unmarked, and plurality is marked by the insertion of an extra *-e-* into the nominal root: *deju* ('group of circles') and *deeja* ('group of squares'). The same process of infixation is used to derive the distinction between the two types of motion: *-e-* is inserted to the verbal form *reeva* ('to bounce') to derive the other verbal form encoding Motion, *reeeva* ('to slide'). There is an apparent functional element *bo* related to the presence of an anchor object. This element is a vestigial marker of aspect from previous generations but in the final language it is not semantically or morphosyntactically conditioned (see examples in main text (7)). Users of the final language CA1 only maintain *bo* to help them produce unique signals with a system which is underspecified for Aspect: participants add it or delete it to provide alternative descriptions when being asked by the computer. The variation within the forms associated with Motion shows another way to satisfy the hard constraint against ambiguity in the signal system whilst maintaining meaning underspecification (i.e., it constitutes meaningless phonemic reduplication).

		Lang CA1	Lang CA2	Lang CA3	Lang CA4
circle	one	deju	fugo	mE^n	pI^n/pU^n
	group	de-e-ju/deju deju	fug-eme	$padamE^n$	$pacupI^n/pacadU^n$
square	one	deja	gopu/gupo	mI^n	dI^n/dU^n
	group	de-e-ja/deja deja	god-eme	$fujI^n$	$dacadI^n/dacadU^n$

Table 2.7: Nominals in the languages evolved in Experiment 3 (from the final generation in all except in CA2, which is taken from the penultimate generation). In languages CA3 and CA4, Nominal elements that referred to the same object shared a root but (inconsistently) differed on the number of repetitions of the last vowel within the root (capitalised vowels with superscripts are used to illustrate this phenomenon). In language CA1 fragments in bold highlight nominal category markers.

		Lang CA1	Lag CA2
slide	ongoing	F ree-e(e)-va (bo) A	tube F tube A
	terminated	F ree-e(e)-va (bo) A	tube F male A
bounce	ongoing	F reeva (bo) A	brillo F brillo A
	terminated	F reeva (bo) A	brillo F male A

Table 2.8: Verbal elements (in context) evolved in the final languages in Experiment 3 (last and penultimate for CA1 and CA2 respectively). In language CA1 fragments in bold highlight verbal category markers.

Experiment 3: nominal forms are phonemically similar within each language, a trait common in Experiment 1 (which also included an artificial pressure) but not in Experiment 2. Table 2.8, on the other hand, shows the verbal elements evolved in Experiment 3; I only show languages CA1 and CA2 because these are the only languages in which verbal elements emerged. Languages CA3 and CA4 only comprise four nominal lexical roots and no verbal elements. These languages satisfy the hard constraint against ambiguity by reduplicating the last vowel of the root n times (without conditioning) and/or shifting the nominals in a description (in transitive motion events). This type of reduplication in languages CA3–CA4 is meaningless, unlike the instances of root-vowel or full morphological reduplication in CA1 and CA2 respectively, where it marks the plurality in nominals, and aspect in verbal forms (see Tables 2.7 and 2.8). Although languages CA3 and CA4 are superficially expressive (each motion event scene is associated with a unique description), most distinctions amongst the meanings expressed are hardly interpretable. Nevertheless, participants manage to communicate with each other quite accurately (see Figure 2.18b). As in Experiment 2, at each matching trial in a communicative event, foils in discrimination arrays are selected at random; most of the time, encoding shape and number of focal and anchor objects is enough for successful communication. Figure

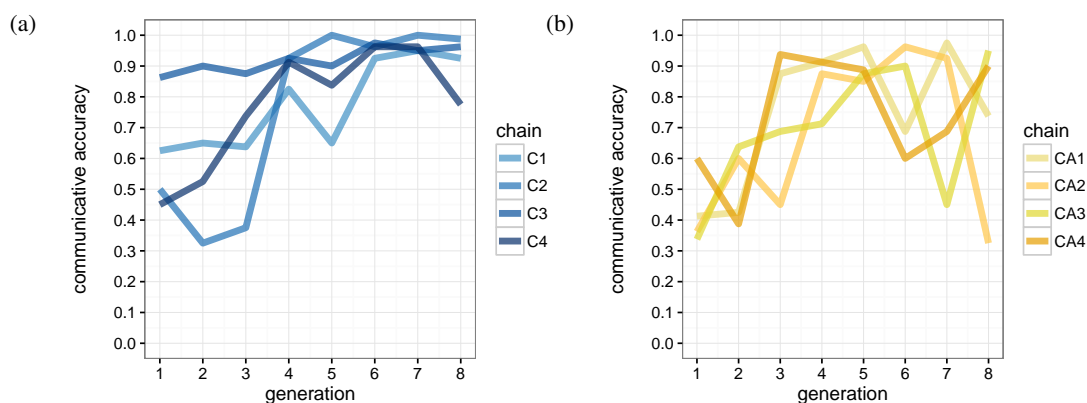


Figure 2.18: This graph shows the communicative accuracy as a proportion of the successes during communication (80 trials) between pairs of participants in Experiment 2 (a) and Experiment 3 (b). We observe an increase in communicative accuracy in the first 5 generations across experiments; however, accuracy decreases in the last generations of Experiment 3, whilst it continued on increasing in Experiment 2.

2.18 visualises communicative accuracy in Experiment 3 (2.18b) compared to Experiment 2 (2.18a): we observe that communicative accuracy increases in the first five generations, but nevertheless, unlike in Experiment 2, it decreases in the last generations on average.

In languages CA1 and CA2 focal arguments tend to precede anchor arguments and in languages CA3 and CA4, which are less systematic in general, word order is free.

2.4.2.2 Compositional structure

Figure 2.19a shows the structure scores obtained from the experimental data in Experiment 3. As in Experiments 1 and 2, I performed a segmented linear mixed-effects model with a breakpoint in generation 4 to explore the effect of generation on linguistic structure across experiments. We will call this Model 7. I used the same fixed and random effects structure as in Model 4, and, as for all models to follow, the fixed effect Experiment was also reverse Helmert coded: Experiment 2 is compared to Experiment 1 and Experiment 3 is compared to the mean of Experiments 1 and 2. The overall model fit was $R^2_{\text{marginal}} = 0.622$ and $R^2_{\text{conditional}} = 0.675$. Figure 2.19b shows the fixed and random estimates obtained in Model 4. The model intercept indicates that languages became significantly structured within the first half of the transmission chains ($\beta = 19.689, SE = 1.931, p < 0.001$). I found significant effects of Generation ($\beta = -5.318, SE = 0.57, p < 0.001$), suggesting that structure increases significantly with each generation in the first 4 generations across conditions. I found a significant effect

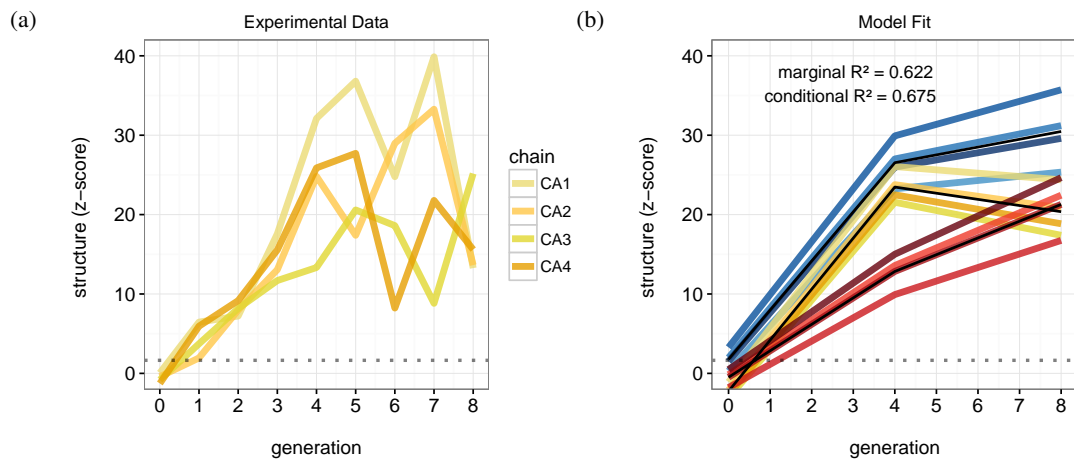


Figure 2.19: (a) Linguistic structure over generations for each of the four transmission chains in Experiment 3. Linguistic structure increases initially as languages are transmitted through generations of learners; however, unlike in Experiments 1 and 2, it seems to start decreasing after generation 4. (b) Fitted values from the mixed-effects Model 7 for Experiment 1 (red), Experiment 2 (blue), and Experiment 3 (yellow). Coloured lines represent the random slopes estimates (for generation) depending on random intercepts (individual chains), whereas the black lines represent the fixed effects estimates for each experiment. In both plots, the dotted horizontal line represents the chance level ($z\text{-score } 95\%CI = \pm 1.645$, one-tailed).

of the interaction between Generation and Experiment 2 ($\beta = 1.431, SE = 0.700, p = 0.044$) but not between Generation and Experiment 3 ($\beta = 0.555, SE = 0.404, p = 0.173$), suggesting that structure increases more rapidly in Experiment 2 than in Experiment 1 (as seen in the earlier analysis) and the slope fit for Experiment 3 is not significantly different from the average of Experiments 1 and 2. After generation 4, slopes significantly flattened across experiments: I found a significant effect of the interaction between Generation and Indicator ($\beta = -4.545, SE = 1.002, p < 0.001$), and no significant effects of the three way interactions with Experiment (Experiment 2, $\beta = 1.9945, SE = 1.2272, p = 0.107$; Experiment 3, $\beta = -1.32, SE = 0.708, p = 0.064$).

I also found a significant effect of Experiment 2 ($\beta = 6.849, SE = 1.831, p < 0.001$) ratifying that languages are more structured at generation 4 in Experiment 2 than in Experiment 1. I did not find a significant main effect of Experiment 3 ($\beta = 1.260, SE = 1.057, p = 0.239$), suggesting that the average structure at generation 4 in Experiment 3 is not significantly different from the average of Experiments 1 and 2.

	gen 1	gen 2	gen 3	gen 4	gen 5	gen 6	gen 7	gen 8
chain CA1	1	0.98	1	1	1	1	1	1
chain CA2	na	1	1	1	1	1	1	1
chain CA3	1	1	1	1	1	1	1	1
chain CA4	1	1	1	1	1	1	1	1

Table 2.9: Proportion of adjacent Shape and Number morphology for a specific object, either focal or anchor. As in the previous experiments, Shape and Number are always encoded adjacent to each other.

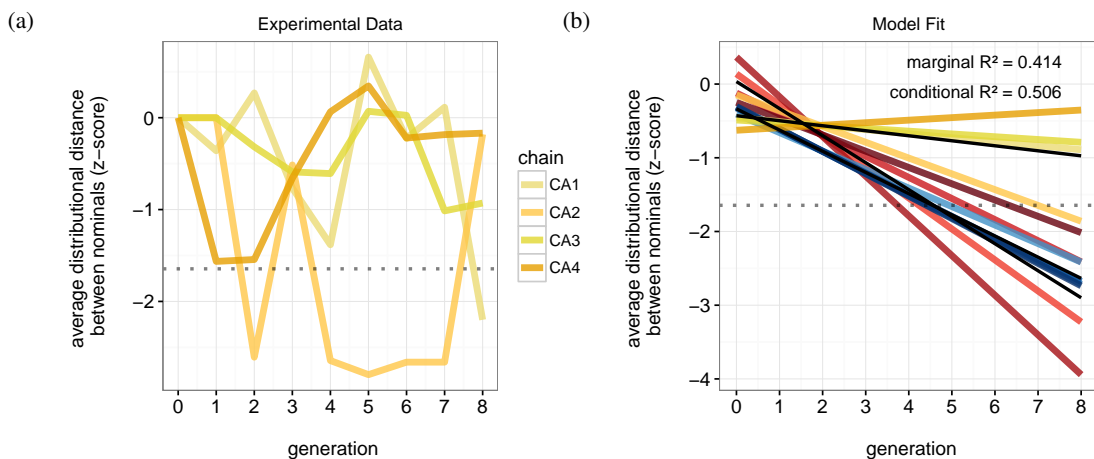


Figure 2.20: (a) Distributional distance between a language's nominals through generations for each of the four chains in Experiment 3. The dotted line represents the chance level (z -score 95%CI = ± 1.645 , one-tailed); z -scores below it indicate that the distributional similarity between nominals is unlikely to arise by chance. (b) Fitted values from the mixed-effects regression Model 8 for the four transmission chains in Experiment 1 (red) and the four transmission chains in Experiment 2 (blue). Coloured lines represent the random slopes estimates (for generation) depending on random intercepts (individual chains), whereas the black lines represent the fixed effects estimates for each experiment.

2.4.2.3 Hierarchical constituent structure: the emergence of complex nominal constituents within sentences

Table 2.9 shows the percentage of times morphs encoding Shape and Number form a continuous constituent. As described in section 2.4.2.1, three languages use suppletive forms for singular and plural objects and only one uses infixation to mark plurality; unlike in previous experiments, it is hard to analyse nominals in Experiment 3 as complex constituents syntactically. Nevertheless, as in the previous experiments, I tested whether these nominals in fact constitute a syntactic category. Figure 2.20a shows the z-scores of the average distance between the distribution of nominals in a language; z-scores below -1.645 indicate that the distributional similarity between nominals is unlikely to occur by chance. Languages in Experiment 3 mostly obtain z-scores within chance level and thus I conclude that, on average, nominal syntactic categories in Experiment 3 did not evolve. We performed a linear mixed effects model, which I will call Model 8, to compare the effect of generation on the distributional distance across experiments. I used the same model structure as in Model 5, and the fixed effect Experiment was reverse Helmert coded as per Model 7: Experiment 2 is compared to Experiment 1 and Experiment 3 is compared to the average of those. The overall model fit was $R^2_{\text{marginal}} = 0.414$ and $R^2_{\text{conditional}} = 0.506$. Figure 2.20b shows the fitted values of Model 8 for fixed and random effects. Model 8's intercept suggests that we do not find a nominal syntactic category by generation 5 on average. The distributional distance between nominals at the intercept was significantly greater in Experiment 3 than in the other two experiments ($\beta = 0.253, SE = 0.076, p = 0.004$). Results show also a significant effect of Generation ($\beta = -0.241, SE = 0.046, p < 0.001$) and a significant interaction between Generation and Experiment 3 ($\beta = 0.086, SE = 0.032, p = 0.019$), suggesting that the decrease in the distributional distance of nominals in Experiment 3 is greater than that in the other two experiments on average (see Figure 2.20b). As in Model 5, I did not find a significant effect of Experiment 2 ($\beta = -0.027, SE = 0.131, p = 0.84$) or a significant interaction between Generation and Experiment 2 ($\beta = 0.040, SE = 0.056, p = 0.492$), ratifying that distributional distance within nominal decreases at a similar rate and to similar degrees. Altogether, these results show that, unlike in previous experiments, complex nominal constituents that form a syntactic category do not evolve in Experiment 3.

2.4.2.4 Word order rules for nominal arguments

The stacked area graphs in Figure 2.21 show the proportions of focal-anchor, anchor-focal and undefined orders of nominal arguments at each generation for each of the four transmission chains in Experiment 3. We observe that the proportion of undefined order only decreases notably in languages CA1 and CA2 as they are transmitted through generations; in the remaining two languages CA3 and CA4, word order is mostly undefined throughout.

I ran a linear mixed effects model, which I will call Model 9, to compare across experiments the effect of generation on the variability of nominal argument orders in a language—calculated by the entropy of the system of orders as described in section 2.2.2.4. The fixed and random effects structure was identical to Model 6 and the fixed effect Experiment was also reverse Helmert coded. Figure 2.22 shows the nominal order variability scores of the experimental data (Figure 2.22a) as well as the fitted values of Model 9 for fixed and random effects (Figure 2.22b). The overall model fit was $R^2_{\text{marginal}} = 0.364$ and $R^2_{\text{conditional}} = 0.415$. Results show a significant effect of Generation ($\beta = -0.054, SE = 0.0106, p < 0.001$) and a significant interaction between Generation and Experiment 3 ($\beta = 0.0193, SE = 0.007, p = 0.015$), suggesting that the order of nominal arguments did not become more consistent with generation in Experiment 3 as it did in the other two experiments on average. Moreover, I found that whilst word order was less variable at the intercept (halfway across the transmission chain) in Experiment 2 than in Experiment 1 ($\beta = -0.09, SE = 0.034, p = 0.02$), it was more variable in Experiment 3 ($\beta = 0.052, SE = 0.02, p = 0.019$).

2.4.3 Discussion

The present study examined whether and how complex compositional structure evolves in the laboratory by increasing the complexity of the meaning space of the languages to be transmitted. I presented three experiments where I manipulated the nature of a pressure for expressivity that interacts with a constant learnability pressure provided by a bottleneck in intergenerational transmission.

Results suggest that linguistic structure evolved to varying degrees between experiments. In Experiments 1 and 2 we showed that compositional constituent structure and positional compositionality evolve as holistic languages are transmitted through generations of participants.

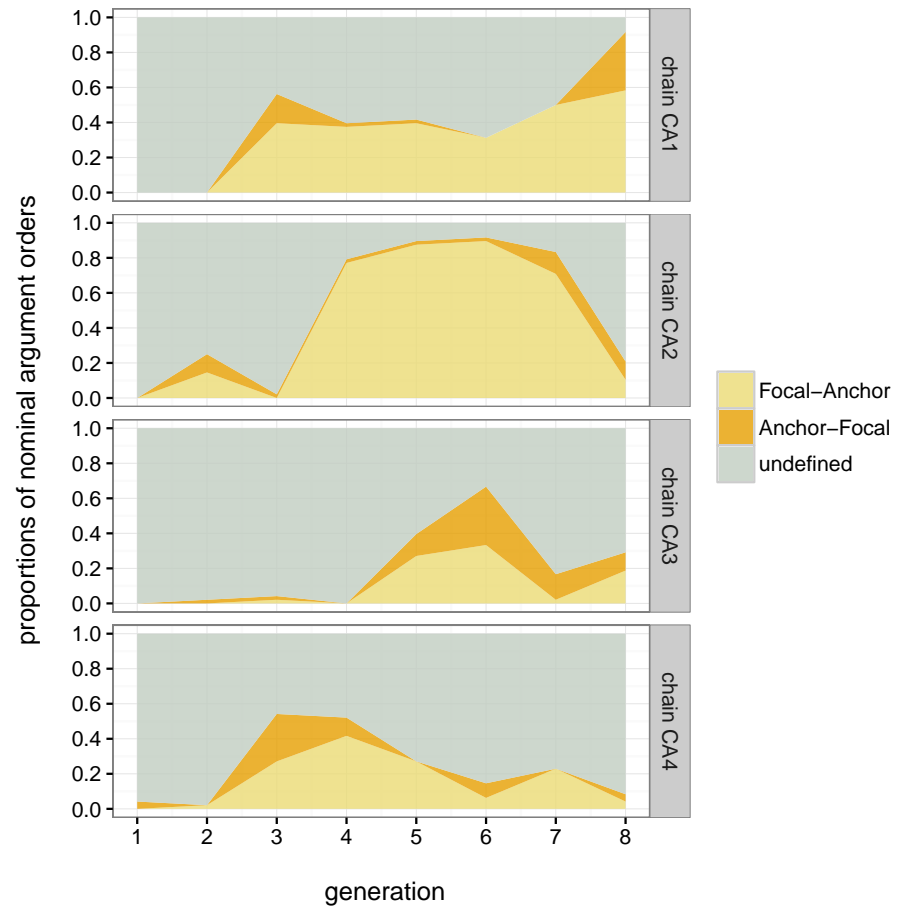


Figure 2.21: Proportion of word order types for nominal arguments by chain and generation. Unlike in Experiments 1 and 2, word order is not fixed in any of the languages that evolved. Only in C2 does word order seem to be fairly systematically Anchor-Focal from generation 4 to 7. In the other chains, word order is mainly undefined (CA3 and CA4) or free (CA1).

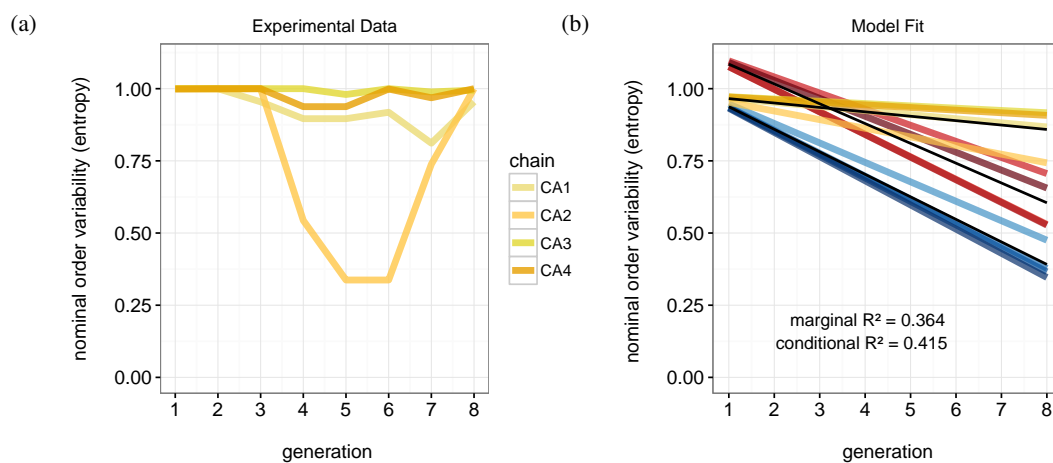


Figure 2.22: (a) Variability of nominal argument orders by generation and chain in Experiment 3. (b) Fitted values from the mixed-effects regression Model 9 for the four transmission chains in Experiment 1 (red), the four transmission chains in Experiment 2 (blue), and those in Experiment 3 (yellow). Coloured lines represent the random slopes estimates (for generation) depending on random intercepts (individual chains), whereas the black lines represent the fixed effects estimates for each experiment.

However, whereas these aspects of compositional structure evolved consistently across languages in Experiment 2, in Experiment 1 they only developed fully in half of the languages. These results suggest that the inclusion of communicative interaction in transmission chains—instead of an artificial pressure against ambiguity—speeds up the emergence of structure, potentially via coordination which facilitates the establishment of linguistic conventions during production. However, without the presence of an artificial pressure against ambiguity, most languages in Experiment 2 were underspecified: i.e., either motion or aspect features were never encoded. With the addition of an extra expressivity pressure in Experiment 3, I expected more expressive as well as complex languages to emerge.

Results from Experiments 3 show that, as predicted, the initial evolutionary rates of structure in Experiment 3 were similar to the average of Experiment 2 and Experiment 1. In other words, structure levels were higher in the first generations in Experiments 2 and 3, than in Experiment 1 and highest in Experiment 2. However, my prediction about later generations was not met: structure scores are significantly lower in Experiment 3 than in Experiment 1. Final languages are less systematic than those obtained in Experiment 2 and even in Experiment 1. These results suggest that the mixture of communication and an artificial restrictive pressure against ambiguity hinders the evolution of linguistic structure eventually. As a result, whilst complex nominal constituents which form syntactic categories can be extracted across

languages in Experiment 2 and in some languages in Experiment 1, they hardly emerge in Experiment 3. Accordingly, no word order rules for nominal constituents evolve in Experiment 3 but they do appear across languages in Experiment 2 and in half of the languages in Experiment 1.

Results from Experiments 2 and 3 thus further support the conclusions from previous studies which suggest that the addition of communication to intergenerational transmission speeds up the emergence of linguistic structure (Carr et al. 2016; Theisen-White et al. 2011). Linguistic conventions are indeed established primarily during interaction (see also Fay & Ellison 2013; Garrod & Anderson 1987; Garrod et al. 2007; Kemp & Regier 2012; Pickering & Garrod 2004; Selten & Warglien 2007). But why does the addition of an artificial pressure against ambiguity into the communicative process hinder the evolution of structure eventually? I propose it is the result of the combination of at least two factors. Firstly, the imposition of expressivity by the artificial pressure hinders establishing conventions incrementally; i.e., the reuse of the same description initially might help interlocutors confirm their hypothesis about the communicative pacts, but the artificial pressure does not allow it (Figure 2.23 shows the proportion of trials in which participants introduced homonymous descriptions before altering after the computer asked them to). Consequently, interlocutors might initially resort to strategies to overcome this hard constraint which, although they might be systematic, would not be compositional: for instance, they might opt to reduplicate characters of lexical items whose reference they want to confirm with their partner (see 2.4.2.1). Secondly, interlocutors establish pacts about these strategies and stick to them; later generations then learn such strategies to discriminate meanings which do not systematically map to any meaning specifically—e.g., reduplication of root vowels an undefined number of times—and perpetuate the strategy. The combination of these factors leads to systematic strategies for discrimination of strings which are easy to learn without the need of structure

In sum, results from Experiment 3 suggests that including an artificial pressure against ambiguity during communicative interaction does not help participants to be more expressive as we expected but instead hinders natural processes in coordination such as the incremental establishment of communicative pacts between interlocutors, ultimately leading to unstructured languages.

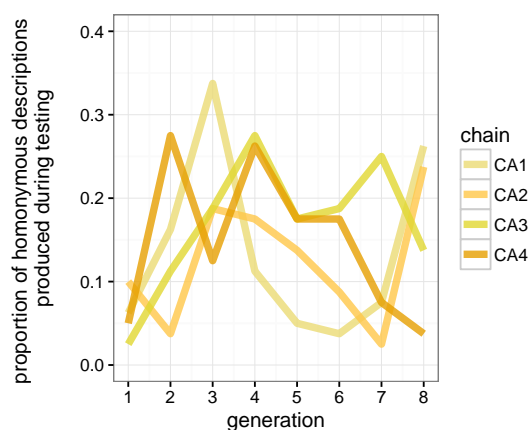


Figure 2.23: Proportions of homonymous descriptions introduced by pairs of participants during testing in Experiment 3, and thus the proportions of trials (out of 80) in which participants were asked to provide an alternative description as they had used it previously to refer to a different scene. We observe an increase of homonymy in the first generations. Later on in chains CA3 and CA4 the proportion of homonyms introduced decreases; however, the contrary tendency is found in chains CA1 and CA2.

2.5 Conclusion

In the experiments comprised in this chapter I show that complex compositional structure which mirrors that found in real languages emerges from cultural transmission. In particular, I demonstrate that the iterated learning model can account for the evolution of linguistic structure beyond basic compositionality: by increasing the complexity of the meaning space, I attested the emergence of hierarchical constituent compositional structure and word order rules for argument marking. This chapter thus provides support for the claim that cultural transmission is a linking mechanism by which the advantages of compositional hierarchical structure (i.e., prerequisite for learnable productive and productively interpretable languages) can permeate language.

Moreover, the work discussed in this chapter demonstrates that both expressivity and coordination pressures provided by communication play an important role in the evolution of linguistic structure. Results from Experiment 1 and 2 suggest that the effect of communication cannot be reduced to an expressivity pressure alone (Garrod & Anderson 1987; Lewis 1968; Pickering & Garrod 2004, 2006; Winters 2017); the need to coordinate during communication facilitates the establishment of linguistic conventions and ultimately—in interaction with learnability and expressivity pressures—speeds up the evolution of complex compositional structure. Additionally, results from Experiment 3 suggest that an extra pressure for

expressivity not introduced naturally into the communicative context trumps the establishment of linguistic conventions hindering the evolution of structure.

2.6 Supplementary analyses

In this section I present measures analysing the learnability and complexity of the systems that evolved to support the claim that languages change over cultural time to become more learnable and to shed further light onto the changes in the complexity of the systems that accompany such an increase in learnability.

2.6.1 Learning error

In order to show that the increase in the compositional structure shown in the experiments above is shaped by the need for language to be learnable and thus favours the learnability of the languages, in this section I show that languages' learnability does indeed increase by generation.

Following the measures used in Kirby et al. (2008, 2015) to evaluate learnability, I define and increase in learnability by the decrease of learning error from the learned system to the produced system. In order to quantify the learning error at each generation, I computed the average normalised Levenshtein edit-distance (LD) (Levenshtein 1966) between the descriptions produced at generation g and those produced at generation $g-1$ to refer to the same scenes; I normalised the distances such that the maximum error is 1. Figure 2.24a shows the learning error in Experiments 1, 2 and 3. In Experiments 1 and 2 we observe a continuous decrease in learning error across generations; however, in Experiment 2, this decrease is greater in the first generations and learning error is lower by the final generation. On the other hand, results from Experiment 3 show a distinct trajectory: we observe a rapid decrease of learning error in the first generations, but it starts to increase after generation 5, along the decrease in structure observed in Figure 2.19.

I performed a segmented linear mixed-effects model with a breakpoint at generation 5 to explore the effect of generation on learning error across experiments. I will call this Model 10. The fixed and random effects structure was the same as in Model 7. The overall model fit was $R^2_{\text{marginal}} = 0.776$ and $R^2_{\text{conditional}} = 0.832$. Figure 2.24b shows the fixed and random estimates

obtained in Model 10. I found a significant effect of Generation ($\beta = -0.111, SE = 0.008, p < 0.001$) suggesting that learning error decreased significantly by generation in the first half of the transmission chains. Results also suggest that it decreased significantly more in Experiment 2 than in Experiment 1 ($\beta = -0.030, SE = 0.010, p = 0.003$), and that the slope in Experiment 3 is similar to the average between the other two conditions ($\beta = -0.007, SE = 0.006, p = 0.251$). After generation 5, the drop in learning error reduces significantly across conditions, more in Experiment 2 than in Experiment 1, and even more in Experiment 3 (grand mean, $\beta = 0.098, SE = 0.0164, p < 0.001$; Experiment 2 compared to Experiment 1, $\beta = 0.047, SE = 0.020, p = 0.21$; and Experiment 3 compared to the average of the other two experiments, $\beta = 0.029, SE = 0.011, p = 0.015$). As observed in Figure 2.24, although the rate of decrease in learning error abates across experiments towards the last generations, in Experiment 3 this is not due to languages stabilising for learning error increases across chains after generation 5.

2.6.2 Paradigmatic and syntagmatic complexity

I calculated the paradigmatic and syntagmatic complexity of languages in terms of Shannon entropy. Paradigmatic entropy measures how variable linguistic segments are within a system and syntagmatic entropy measures how variable meaningful linguistic segments are within a given description. As compositional structure evolves, morphs emerge to map to different features of the meaning space establishing isomorphism between syntax and semantics. Consequently, the monomorphic descriptions comprised in the initial holistic languages become polymorphic and each morph is a segment that is reused often within a system. I thus expect paradigmatic entropy to decrease and syntagmatic entropy to increase as holistic languages becomes compositional.

I obtained productive linguistic segments automatically by extracting all matching segments between words in a language (i.e., strings of characters typed between spaces within descriptions). For example, from a language that contains the set of words $\{evo, roji, ref, tube, babatube\}$, I would have obtained the following set of linguistic segments types: $\{evo, roji, ref, tube, baba\}$. I then obtained the paradigmatic entropy of a language by calculating the Shannon entropy of the whole linguistic system given the extracted set of segments. More specifically, to quantify the paradigmatic entropy of language X I calculated $H(X) = - \sum_{i=1}^n P(x_i) \log_2 P(x_i)$, where the sum is over the different segments in the set and $P(x_i)$ is the relative frequency

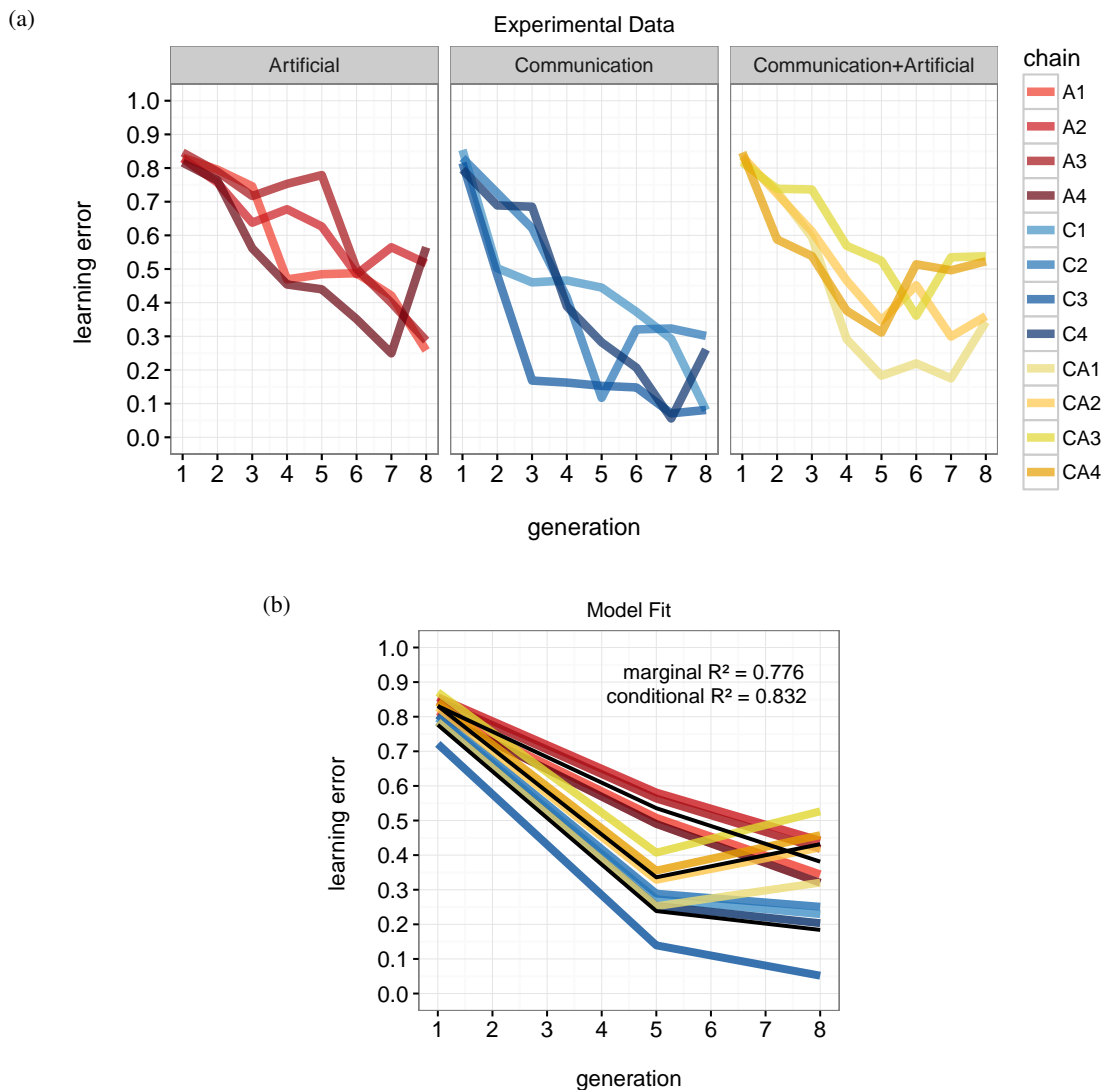


Figure 2.24: (a) Learning error over generations across experiments for each of the transmission chains. Whilst in Experiment 1 (Artificial) and Experiment 2 (Communication) learning error decreases consistently as languages are transmitted through generations of learners, in Experiment 3 (Communication + Artificial) learning error starts to increase after generation 5—in accordance with the evolution of structure (see Figure 2.19). (b) Fitted values from the mixed-effects Model 10 for Experiments 1, 2, and 3. Coloured lines represent the random slopes estimates (for generation) depending on random intercepts (individual chains), whereas the black lines represent the fixed effects estimates for each experiment.

of segment x_i in language X . By contrast, in the calculation of syntagmatic entropy I am not interested in mere signal complexity within descriptions but in syntagmatic *morphological* complexity; consequently, and unlike for paradigmatic entropy, I only take into account linguistic segments that can be considered morphs, i.e., productive segments whose mapping to a meaning feature values is significant. To calculate syntagmatic entropy I first computed the entropy of each description individually (80 in total in each language) given the extracted set of morphs. More specifically, to quantify the syntagmatic entropy of a sentence S , I first calculated $H(S) = -\sum_{i=1}^n P(s_i) \log_2 P(s_i)$, where the sum is over the different morphs in the set (if meaningful), and $P(s_i)$ is the relative frequency of morph s_i in sentence S . I then took their average as the syntagmatic entropy score for the language as a whole.

Figure 2.25a shows the paradigmatic entropies of the languages as they evolve: paradigmatic entropy decreases with generation across experiments. A visual inspection reveals that paradigmatic entropy decreases faster in the experiments that include communication (i.e., Experiments 2 and 3) than in Experiment 1 and we observe the most abrupt decrease in paradigmatic entropy in Experiment 2. I ran a growth curve analysis (GCA) to explore the relationship between paradigmatic complexity and generation. I utilised GCA because I do not expect to see the hypothesised effects equally pronounced across generations. Instead, I expect them to be more distinct in the first generations and less so in the latter ones as languages become more stable as a result of the cumulative increase in learnability. Moreover, in Experiment 2, where underspecification is permitted (i.e., there is no hard constraint against ambiguity), if conventions are established incrementally, I expect paradigmatic complexity to drop dramatically in the first generations and to slightly increase in the later generations as more form-meaning mappings are settled. GCA thus allows me to capture the expected trajectories more accurately than the SLMM used in previous analysis.

As fixed effects in this Model 11, I introduced a linear predictor for Generation (centred) as well as its interaction with Experiment (Experiment 1, 2 and 3, reversed Helmert coded in that order). I also added a quadratic term, Generation^2 to allow us to explore the change in the effect of Generation as we move further along in the transmission chain. As random effects I introduced an intercept for Chain as well as by-Chain slopes for the effect of Generation and Generation^2 . The overall model fit was $R^2_{\text{marginal}} = 0.747$ and $R^2_{\text{conditional}} = 0.840$; model estimates are visualised in Figure 2.25b. I found a significant effect of Generation

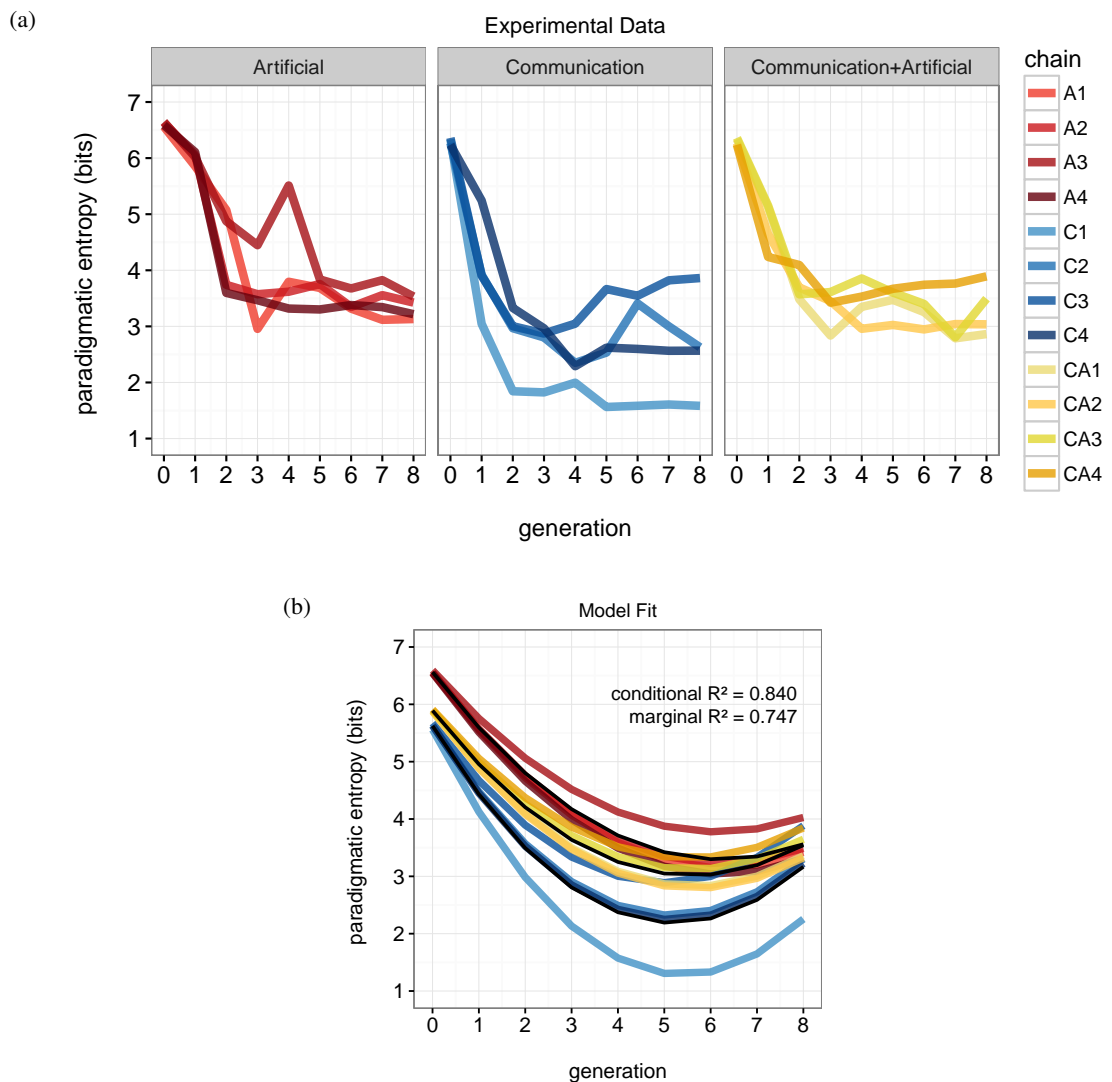


Figure 2.25: (a) Paradigmatic entropy over generations for each transmission chain. Chains in Experiment 1 (Artificial), Experiment 2 (Communication), and Experiment 3 (Communication + Artificial) are coloured in red, blue and yellow respectively. Paradigmatic entropy decreases as languages are transmitted over generations. (b) Fitted values from the mixed-effects regression. Coloured lines represent the random slopes estimates (for generation) depending on random intercepts (individual chains), whereas the black lines represent the fixed effects estimates for each experiment.

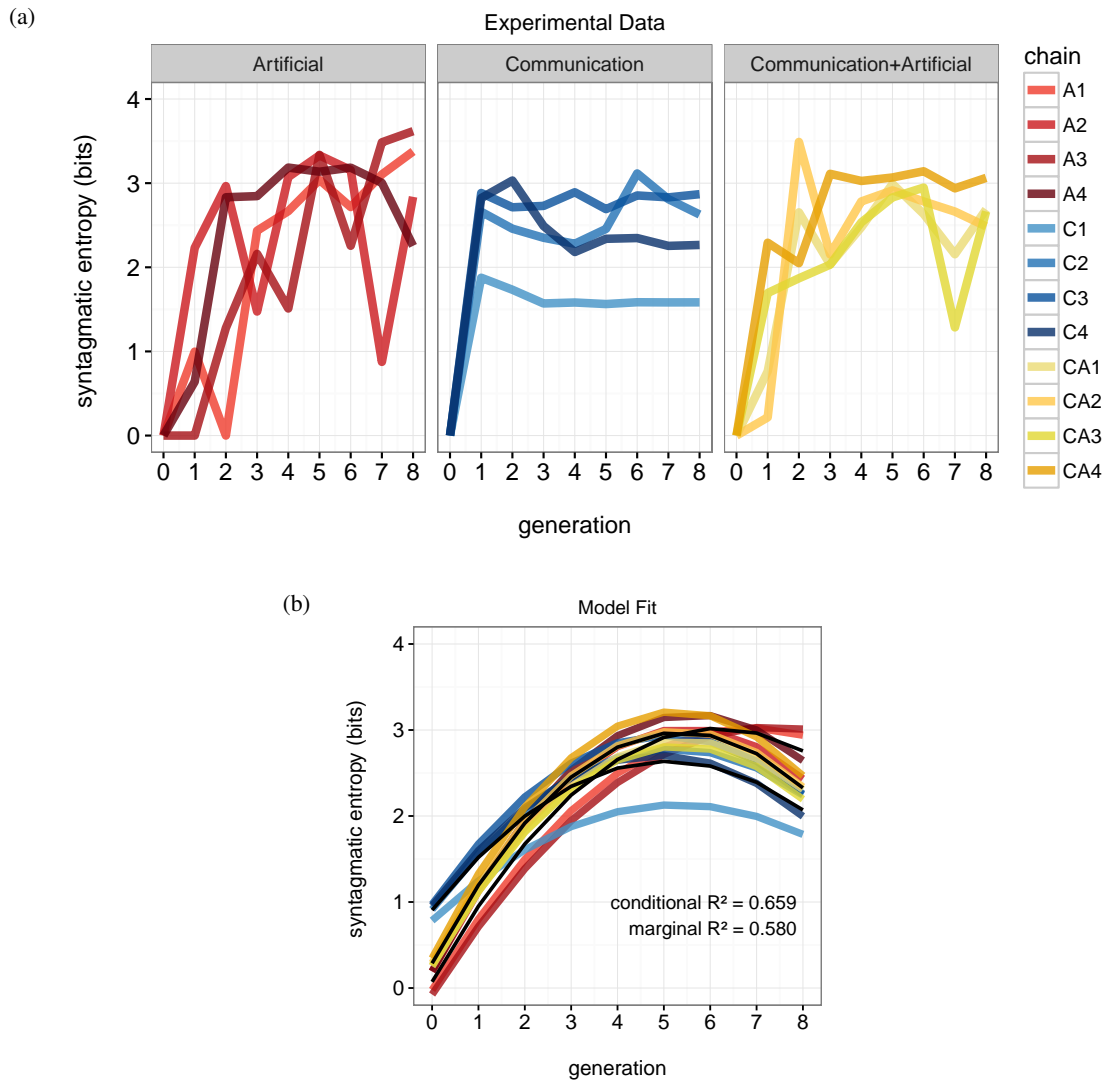


Figure 2.26: (a) Syntagmatic entropy over generations for each transmission chain. Chains in Experiment 1 (Artificial), Experiment 2 (Communication), and Experiment 3 (Communication + Artificial) are coloured in red, blue and yellow respectively. Syntagmatic entropy decreases as languages are transmitted over generations. (b) Fitted values from the mixed-effects regression. Coloured lines represent the random slopes estimates (for generation) depending on random intercepts (individual chains), whereas the black lines represent the fixed effects estimates for each experiment.

($\beta = -0.325, SE = 0.0256, p < 0.001$) suggesting that paradigmatic entropy decreases significantly over time. However, the slope does not remain unchanged as we move further along in the transmission chain: the significant effect of Generation² ($\beta = 0.101, SE = 0.009, p < 0.001$) suggests that the drop in paradigmatic entropy by generation reduces significantly. Results suggest no differences in the effects of Generation and Generation² across experiments (lowest: $\beta = 0.0341, SE = 0.031, p = 0.298$) aside from a near-marginal effect of the interaction between Generation² and Experiment 2 ($\beta = 0.021, SE = 0.011, p = 0.069$), suggesting that the change in slope might be slightly, although not significantly, greater in Experiment 2. The data obtained for Experiment 2 shown in Figure 2.25 suggest that the drop in entropy is abrupt in the first generation and remains relatively constant from generation 2 onwards, with a slight increase towards the end. Results do show a significant effect of Experiment 2, suggesting that paradigmatic entropy is lower in Experiment 2 than in the Experiment 1 at generation 4 ($\beta = -0.670, SE = 0.179, p = 0.004$). Entropy scores at the same generation in Experiment 3 are similar to the average between Experiment 1 and Experiment 2 ($\beta = 0.069, SE = 0.104, p = 0.521$). In sum, paradigmatic entropy decreases with generation and more so in the first than in the latter generations. Moreover, the entropy drop is most abrupt in Experiment 2.

On the other hand, Figure 2.26a shows the syntagmatic entropy of languages as they are transmitted down the chain and become more structured. Note that syntagmatic entropy is 0 on initial languages because no productive segments are encountered within descriptions as they are monomorphemic—i.e., holistic mappings to motion events scenes. We observe that syntagmatic entropy increases with generation across experiments. However, in Experiment 2 syntagmatic entropy increases abruptly in the first generation and remains constant thereafter. In the other two experiments, the increase seems to be more gradual, particularly in Experiment 1. I ran a growth curve analysis with the same model structure to explore the effect of generation on syntagmatic entropy. As in the previous model, I expect the effect of generation to not be as pronounced towards the final generations as structure and learnability increases. The overall model fit was $R^2_{\text{marginal}} = 0.580$ and $R^2_{\text{conditional}} = 0.659$; model estimates are visualised in Figure 2.26b. The model intercept indicates that syntagmatic entropy is significantly greater than zero at generation 4 on average ($\beta = 2.671, SE = 0.136, p < 0.001$) and thus that descriptions become polymorphemic in the first half of the transmission chain; scores at the intercept

were also similar between experiments (lowest: $\beta = 0.064, SE = 0.096, p = 0.519$). Moreover, results suggest a significant effect of Generation ($\beta = 0.245, SE = 0.028, p < 0.001$) suggesting that syntagmatic entropy increases significantly over cultural time; but the slope abates as we move further along in the transmission chain ($\beta = -0.079, SE = 0.011, p < 0.001$). No differences were found for Generation and Generation² across experiments (lowest: $\beta = 0.007, SE = 0.008, p = 0.407$) aside from a significant interaction between Generation and Experiment 2 ($\beta = -0.095, SE = 0.0343, p = 0.019$), confirming that the effect of generation on syntagmatic entropy in Experiment 2 is not as pronounced —entropy scores stabilise after generation 1 (see Figure 2.26a). In sum, syntagmatic entropy increases relatively abruptly in the first generation across experiments, and whereas it continues on increasing in Experiments 1 and 3, it does not in Experiment 2.

2.6.3 Discussion

The measures presented in this section further support the conclusions drawn throughout the chapter. Results show that languages become more learnable through cultural transmission, thus replicating the findings in previous IALL studies (Beckner, Pierrehumbert, & Hay 2017; Carr et al. 2016; Kirby et al. 2014, 2015; Motamedi et al. under revision). Compositional structure increases the learnability of languages as it allows learners to reproduce all descriptions without the need of a model for every single one. From a reduced set of descriptions generated from a compositional grammar, learners can deduce the lexicon and the grammatical rules required to reproduce the whole language. Consistent with the previously discussed results, learning error decreases by generation across experiments, and more in the first generations than in the latter generations as learning error approaches floor.

However, in Experiment 3, I showed that, consistent with a decrease in structure, learning error increases in the second half of the chains. In section 2.4.3 I discussed that the strategies learners develop to satisfy the hard constraint on expressivity (brought on by the artificial pressure) were very easy to learn. However, I suggested that those strategies for discrimination of strings did not allow for the discrimination of meanings; i.e., languages were superficially expressive but highly degenerate. Therefore, although learners might have learned that discrimination between description is provided via reduplication of specific characters, it is highly unlikely that they learned the exact number of reduplications for each description present in

the input; as for descriptions without a model, it is even less likely that learners can reproduce them.

Results in this section also show that the increase in learnability is parallel to the reduction of paradigmatic entropy. As compositionality emerges from holistic languages, we move from a system with no productive segments to a system with few but very productive segments—i.e., productive morphology. Paradigmatic entropy decreases abruptly in the first generations and it remains fairly stable thereafter. Entropy is lowest in Experiment 2 as a product of two features: most languages are underspecified (i.e., contain fewer morphs that map to meaning features), and words are mainly inflected via total reduplication (i.e., morphs are fewer and even more productive). Moreover, we also observed a slight increase in paradigmatic entropy in the latter generation in Experiment 2. I proposed that this slight increase is supporting evidence for an incremental process of conventionalisation. Accompanying the reduction of paradigmatic entropy, we observed an increase in syntagmatic entropy over generations. As morphology emerged to match the different features of the meaning space, descriptions became polymorphic and thus they included a greater number of different morphs. These results suggest that learnability is facilitated by a low paradigmatic entropy, further supporting the claim that learnability in languages is provided by the compressibility of the finite sources from which learner can make infinite use of: i.e., the morphological lexicon in this measure of paradigmatic entropy, and the grammar in Kirby et al. (2015). It is precisely the productivity achieved by compositional structure that potentially allows the production of syntagmatic complexity otherwise impossible to be reproduced by learners.

In sum, in this section I provide further support to the claim that languages become more learnable as they undergo cultural transmission. Moreover, the results from the paradigmatic and syntagmatic entropies provide further insight into the dynamics of change from holistic languages with monomorphic descriptions to compositional languages with polymorphic descriptions: morphological lexicons are reduced and the morphological complexity that descriptions achieve necessarily increases.

2.7 Summary

The work discussed in this chapter provides support for the role of cultural transmission in the evolution of compositional hierarchical structure—a prerequisite for productive and productively interpretable languages, as well as argument marking via word order rules. Results demonstrate the importance of the combination of communicative interaction and transmission processes in the evolution of linguistic structure as well as the importance of the complexity of the world to express and communicate about. These results corroborate findings from previous iterated learning studies and crucially expand the linguistic complexity that can be obtained from laboratory models of cultural transmission.

Chapter 3

The effect of communicative interaction on efficiency and complexity

3.1 Introduction

Language and communication go hand in hand across human cultures; language is necessarily transmitted through usage in communicative interaction. Learners acquire a language from exposure to its use in a communicative context. Speakers, at the same time, mostly produce language for communicative purposes. It is then crucial to the study of language evolution to pin down the effects on linguistic behaviour of the mechanisms involved in communicative interaction.

In Chapter 2 I highlighted the importance of communication in the evolution of language; in combination with transmission, communicative interaction facilitates the emergence of compositional constituent structure. We observed that languages rapidly evolved compositionality and simple lexicons as well as highly regular rules of word formation and word order. Without communicative interaction, structure did not evolve as rapidly, suggesting that coordination facilitates the early establishment of linguistic conventions leading to linguistic structure. Moreover, we observed that conventionalisation proceeds in a piecemeal fashion whereby interlocutors learn about each other's linguistic knowledge and gradually align with it; if this

piecemeal process is hindered, systematic conventions are less likely to arise. The work discussed in Chapter 2 thus hinted that this need for learning during coordination in the communicative context might also facilitate the establishment of structured and generalisable linguistic conventions.

Previous work studying the mechanisms involved in interaction claim that communicative contexts are not merely situations of information transfer where coding and decoding of signals takes place (Pickering & Garrod 2004). Interlocutors do not contribute autonomously to interaction; instead, they engage in a joint activity of collaborative problem solving (H. H. Clark & Wilkes-Gibbs 1986; Garrod & Anderson 1987; Pickering & Garrod 2004). Each communicative event presents interlocutors with a coordination problem, where they must use linguistic forms or strategies that agree with one another to achieve mutually acceptable outcomes (Lewis 1969). Communicative success thus requires interlocutors to coordinate and align on a shared system of conventions, whereby interlocutors can be situated in equivalent information states. Previous research has shown that the establishment of conventions allows the maximisation of communicative efficiency and not only effectiveness (H. H. Clark & Schaefer 1989); interlocutors utilise conventions to contribute to more efficient communication by reducing effort in production as well as in comprehension. In which ways does the drive for effort minimisation during interaction affect the complexity of linguistic systems?

Interaction studies have provided sparse evidence for the drive to reduce effort in communication, i.e., by reusing the same expression to refer to the same meaning or by shortening the length of expressions (H. H. Clark & Wilkes-Gibbs 1986; Krauss & Weinheimer 1964, 1966). However, the systematic comparison between communicative and non-communicative context is scarce, making it hard to depict the specific aspects of the complexity of communication systems that can be attributed to actual communicative interaction or to other aspects of linguistic production (e.g., frequency of use or priming, which could also result in the reduction of effort and complexity, see e.g. Kanwal et al. (2017) and Fehér, Wonnacott, and Smith (2016)). In this chapter I experimentally investigate the effect of communicative interaction (without transmission or language learning) on linguistic complexity, allowing speakers to use their native language. Using the same experimental set-up as in Chapter 2, I test differences in the complexity of production systems obtained from individuals in isolation and production systems obtained during dyadic communicative interaction.

3.2 Convention and communicative interaction

Language is not simply an internal system used by individuals in isolation, language is mostly transmitted and used in communicative contexts. At the same time, communication is not merely a process of information transfer, where interlocutors contribute autonomously. In order to communicate effectively interlocutors must align with each other on a shared system of linguistic forms and strategies, i.e., linguistic conventions. Without these conventions, interlocutors will not be able to align on equivalent informational states and communicative success will be extremely difficult to achieve (H. H. Clark 1996; Lewis 1969).

Lewis (1969) proposed that *coordination* should be studied as a distinctive social competence needed to solve the aforementioned coordination problems. Through coordination, interlocutors align on a set of *conventions*, which are the canonical forms that coordination takes when it is grounded in the interlocutors' knowledge and experience of one another's (flexible) behaviour. In this tradition, alignment between interlocutors is thought to arise from the establishment of shared knowledge or common ground facilitated by coordination (H. H. Clark 1996; H. H. Clark & Marshall 1981; Lewis 1969). Importantly, solving coordination problems requires all parties involved to align both in content and process (H. H. Clark & Wilkes-Gibbs 1986; Grice 1957; Lewis 1969). With each utterance, the speaker tries to convey their intended meaning tailored to the hearer; the hearer knows this, and tries to arrive at the correct interpretation. Hence to coordinate by convention entails recognising the intentions of others as well as the utterances (i.e., knowledge on both content and process needs to also be shared).

However, there exists a more mechanistic account of coordination during communicative interaction put forward by Pickering and Garrod (2004) called the *interactive alignment model*, whereby alignment between interlocutors is proposed to arise from low-level automatic priming processes. In this account, the information that is shared between the interlocutors constitutes what Pickering and Garrod (2004) call the implicit common ground. Unlike common ground, implicit common ground does not derive from the interlocutors' explicit knowledge of one another's states but is instead thought to be built up automatically (e.g., through reciprocal priming) and reformulated through "straightforward" processes of repair (also not requiring explicit knowledge) when faulty (for further detail, see Pickering & Garrod 2004). The interactive alignment model suggests that these automatic and low-level processes are enough

for alignment to arise and only when the primitive mechanisms such as priming fail to produce alignment, more sophisticated strategies which depend on representing the interlocutor's mental state (i.e., theory of mind) are required.

Both of the discussed accounts of communicative interaction highlight the need for alignment on a system of linguistic conventions to achieve communicative success, and that alignment is facilitated by the access to shared knowledge or information states as well as by the presence of feedback. Shared knowledge allows interlocutors to produce and interpret form-meaning mappings with higher probability of being mutually accepted based on previous experience. Feedback allows interlocutors to test these hypotheses about form-meaning mappings and modify their behaviour accordingly. Positive feedback (i.e., confirmation) is necessary for interlocutors to know that they understand each other and thus that form-meaning mappings are shared amongst them. Negative feedback (i.e., disagreement) might allow interlocutors to realise they do not understand each other and thus that form-meaning mappings are not shared and need repair; speakers can then reformulate and clarify the intended meaning, and hearers can adjust their form-meaning mappings to align their interpretation with the speaker's intended meaning. Corrective feedback, just as negative feedback, lets the speaker know that the meaning intended might not be interpreted accordingly but provides an alternative form to map to the intended meaning which is considered more appropriate by the hearer; the speaker then can update their form-meaning mappings to align with their partners'.

Communication thus requires interlocutors to coordinate and align on a shared system of conventions (H. H. Clark & Wilkes-Gibbs 1986; Lewis 1969; Pickering & Garrod 2004). However, efficient communication requires that coordination interacts with other pressures at play in language transmission, namely those of expressivity and learnability; i.e., the conventions established need to be both expressive and learnable in order to solve the immediate task at hand as well as future interactions (Winters 2017). Since the communicative history constrains future outcomes (Millikan 1998), if conventions are not generalisable or expressive enough, communicative systems can end up in suboptimal states where they do not allow themselves to be built upon further and eventually render communication of novel meanings more effortful.

3.3 Least collaborative effort and linguistic complexity

H. H. Clark (1996) claims that the collaborative process in which participants engage to align on a shared system of conventions leads also to participants engaging in what the author—expanding Zipf’s principle of least effort (Zipf 1949)—has called *least collaborative effort* (H. H. Clark & Wilkes-Gibbs 1986), i.e., minimising the total effort spent during interaction, in both production and comprehension of utterances (H. H. Clark & Schaefer 1989; Davies 2006). Under this view, interlocutors adjust conventions to exploit the communicative context. Various types of experimental evidence further support this principle and show that the adjustment interlocutors make to the communicative context might affect the structure of the utterances produced (H. H. Clark & Wilkes-Gibbs 1986; Hupet & Chantraine 1992; Krauss & Weinheimer 1964, 1966) as well as the complexity of the overall local (i.e., pertaining to a specific situation) system of communication (Brennan & Clark 1996; Fox Tree 1999; Fussell & Krauss 1989; Garrod & Anderson 1987).

Experimental research looking at the emergence of convention has extensively reported that over the course of interaction, language users shape communicative systems to minimise the effort in production (Atkinson 2016; Brennan & Clark 1996; H. H. Clark & Wilkes-Gibbs 1986; Garrod & Anderson 1987; Krauss & Weinheimer 1964, 1966; McAllister, Potts, Mason, & Marchant 1994; Murfitt & McAllister 2001). In a pioneering experiment, Krauss and Weinheimer (1964) had pairs of participants play a dyadic communication game, where they were asked to repeatedly describe novel objects in English for their partners to identify from a finite set—this paradigm will henceforth be referred to as a repeated-reference task¹. Without previous experience about communicative success, participants could not know how much information was required and thus initially, they would use long expressions to describe the novel object. However, as participants built shared knowledge over repeated communicative events, they would gradually shorten referring expressions to maximise communicative efficiency. One pair of participants, for example, started to describe an image as “upside-down martini glass in a wire stand” and over the course of interaction, the description reduced to

¹Note that unlike other reference tasks in experimental extensions of agent-based models (e.g., iterated learning paradigms as per Kirby et al. 2015 or other signalling games as the naming game Steels 1995), these reference tasks are always in the context of natural language usage and not in the context of establishing a new system of communication. I will thus refrain from discussing the former in this chapter and refer the reader to section 1.3 and Chapter 2 for a discussion of the relevant “reference” tasks in the context of artificial language learning and evolution.

“inverted martini glass”, then “martini glass”, and finally participants converged on “martini”. These results thus show that alongside the conventionalisation of referring expressions, the length of these expressions reduces over repeated interaction as a function of their frequency of use. Critically, later work showed that these results were dependent on participants receiving positive and/or concurrent feedback (Hupet & Chantraine 1992; Krauss & Weinheimer 1966), suggesting that the communicative context is in fact driving the reduction in effort. Krauss and Weinheimer (1966) showed that participants shortened expressions over time more when they received concurrent feedback than when they just received positive feedback after having described an object; without a high percentage of positive and/or concurrent feedback, participants did not shorten descriptions over time. Further support to framing communication as a promoter of the reduction in effort comes from Hupet and Chantraine (1992), who had participants repeatedly describing tangram pictures after having been told either that all their descriptions would reach the same recipient or that they would reach different recipients each time. Participants did not shorten their descriptions over repeated use in either case.

Altogether these results suggest that mutual acceptance of an expression and thus shared knowledge—only provided by the communicative context—is necessary before interlocutors reduce production effort; the shortening of expressions cannot be explained by mere repetition and/or recency of production. Interlocutors attune their productions to the information state of their addressee, only then allowing the attainment of simpler and more effective local systems of communication. Moreover, performance in interaction can be further facilitated by mechanisms in coordination which do not require explicit attuning to the interlocutor’s information state; e.g., alignment achieved from reciprocal priming would also simplify production as well as comprehension during dialogue and altogether minimise communicative effort (Pickering & Garrod 2004). Further experimental studies support these findings (Brennan & Clark 1996; H. H. Clark & Wilkes-Gibbs 1986; Garrod & Anderson 1987) and highlight more features of simplification of systems through repeated usage in interaction: participants converged to less diverse lexicons. Linguistic alignment can be achieved through the mutual acceptance of form-meaning mappings. Once aligned, participants can repeatedly use the same expression to refer to the same meaning with assured communicative success. Interlocutors stick to these conventions in order to ease communication (i.e., retrieval and comprehension) and be able to benefit from a local principle of meaningful contrast (E. V. Clark 1988); when a new term

is used, hearers will rapidly understand there is a new meaning dimension that the speaker is referring to which has not been discussed before. The use of a new term for the same meaning would mislead hearers to think the speaker wants to convey new information (E. V. Clark 1988). Communication thus benefits from interlocutors being conservative in their linguistic choices, which leads to a reduced lexicon once conventions are established. In comparison to isolate production, such conservatism might also result in lower overall linguistic complexity in the set of expressions used during communicative interaction. Michel (2011) provides support for this claim: in argumentation tasks, adult speakers produce less complex lexicons and structures in communicative interaction than in isolate production.

The communicative context also influences how easy referring expressions are to comprehend. Speakers adapt their linguistic behaviour as they learn to match the information the audience requires (and no more), which allows them to increase communicative efficiency (H. H. Clark & Wilkes-Gibbs 1986; Davies 2006; Fox Tree 1999; Fussell & Krauss 1989; Horton & Gerrig 2002). Fussell and Krauss (1989) had participants create (written) referring expressions either for another speaker or for themselves; the same subjects were later asked to match the intended referents for their own expressions and for those of other speakers (who at the same time had encoded expressions intended to another speaker or to themselves). The authors showed that written messages intended for oneself overall contained higher lexical complexity; i.e., they used more diverse lexicons, contained more words of lower frequencies (i.e., type-token ratios were higher) and were more likely to contain figurative descriptions. However, their study also shows that written messages to oneself were shorter than those intended to others, which does not match the results of previously discussed studies on the reduction of description length over repeated use (e.g., see Krauss & Weinheimer 1966). However, in Fussell and Krauss (1989), participants are not involved in explicit communicative interaction or receive any real-time feedback, and thus speakers cannot ground common knowledge and exploit it to increase efficiency in production. Traxler and Gernsbacher (1992) already noted that even the slightest form of feedback helps speakers to envision how hearers interpret written descriptions and gradually maximise efficiency.

Fussell and Krauss (1989) also showed that while all subjects correctly matched referents for the expressions produced by themselves, they also did significantly better with others' messages when they were intended for another speaker. Fox Tree (1999) provided similar re-

sults in the spoken modality: even when intended for others (i.e., a third-party passive listener or over-hearer), instructions that are a product of monologue rather than dialogue are more difficult to follow. This may be because dialogues contain a greater number of perspectives supplied by both the speaker and the addressee, and thus increase the likelihood of there being a perspective which is understood by a third person (Fox Tree & Mayer 2008). It might also be due to the shared-knowledge grounding process (especially through the feedback provided by the addressee about their level of understanding), which may increase the likelihood that the descriptions will be comprehensible for any individual, not just those directly involved in the interaction (Branigan, Catchpole, & Pickering 2011). Moreover, over-hearers (who passively participate in dialogue) understand references less well than intended addressees; the latter appear to have an advantage over over-hearers because addressees actively participate in the communicative process (Schober & Clark 1989) and guide the developing descriptions to those they prefer to adopt (Branigan et al. 2011). These results are consistent with a shared knowledge framework: since shared knowledge with oneself is maximal and no attuning to listeners is required, there is not as much cost in producing idiosyncratic expressions (i.e., communication to oneself is not hampered), but their comprehension by another speaker could be negatively affected. These results also suggest that the mechanisms involved in attuning to the audience are necessarily different from the automatic priming which interlocutors are said to primarily resort to in the interactive alignment model proposed by Pickering and Garrod (2004); i.e., self-priming is not playing a role and reciprocal priming cannot take place, nevertheless, interlocutors adapt to their audience to minimise comprehension effort, highlighting the role of more sophisticated mechanisms (e.g., theory of mind).

3.4 A note on language learning and communicative interaction

Given the need for coordination in communicative interaction (which requires interlocutors to learn about each other's knowledge), communication has often been framed as a learning context (H. H. Clark & Wilkes-Gibbs 1986; Horton & Gerrig 2002; Long 1985; Michel 2011). Undoubtedly, interactive communication provides the learner with learning opportunities. Moreover, interaction heightens attention to meaning and form as it requires interlocutors to engage in a collaborative problem-solving task with a joint focus, whereby, in addition, they

are provided with feedback from one another on how well they are conveying information.

As a matter of fact, first language acquisition research demonstrates that both joint focus and feedback facilitate language learning. Since Bruner and Watson (1983), who postulated a link between the presence of joint focus in communicative interaction and the development of reference, many longitudinal studies have indeed found evidence for a positive correlation between joint focus in child-parent interactions and a child's lexical development (Markus, Mundy, Morales, Delgado, & Yale 2000; C. B. Smith, Adamson, & Bakeman 1988; Tomasello & Farrar 1986) (for a critical review of the literature, see Akhtar & Gernsbacher 2007). On the other hand, it is obvious that without positive feedback, a naive learner would never be able to know whether they communicate appropriately. Not so obvious or at least not as uncontroversial has been the effect of corrective feedback on language acquisition. Corrective feedback is not as present in a child's input as positive feedback; however, many studies suggest that the existing negative feedback does in fact facilitate linguistic competence (Farrar 1992; Scherer & Olswang 1984; Strapp, Bleakney, Helmick, & Tonkovich 2008) (for a critical review, see Schoneberger 2010).

Second language acquisition research has also highlighted the importance of interaction for interlanguage development (Long 1985). During interaction learners are naturally pressured to make meaningful use of their linguistic knowledge because they need to understand and be understood. Moreover, learners engaged in interaction are involved in negotiations of form and meaning and learn through hypothesis testing; when they fail at being comprehended they receive negative feedback from their communication partner. Support for the importance of interaction in L2-learning is given in Michel (2011); the author shows that L2-learners produce more accurate and fluent speech when they are engaged in a dialogue rather than in isolate production. Isolate production provides fewer opportunities for language learning; due to the lack of feedback, isolate production does not require learners to pay as much attention to form and meaning.

Studies looking at production from adult L1-speakers also point to aspects of the communicative context that ease linguistic performance which echo those aspects highlighted to help learners acquire a new language. Michel (2011) demonstrates that L1-speakers are also more accurate and fluent in dialogue than in isolate production. These results suggest that aspects of the communicative context which have shown to facilitate language acquisition are pervasive in

communicative interaction with adult L1-speakers. Altogether, these results show that mechanisms that allow successful and efficient communication, which tend to draw upon simpler and (in some aspects) more conservative linguistic systems than those produced in isolation, help language learning.

3.5 Experiment 4

The literature reviewed above suggests that communicative interaction should be regarded as a joint activity akin to solving coordination problems in order to reach effective and efficient communication. In order to solve coordination problems, interlocutors must align on a shared system of conventions. Alignment draws upon at least two important features provided by the communicative context: shared knowledge as well as feedback. Interlocutors exploit the affordances of the communicative context for communicative efficiency, i.e., minimising the total effort spent by interlocutors to achieve successful communication during dialogue. Maximisation of efficiency benefits from systems which are generalisable and easy to produce as well as easy to understand, altogether minimising effort at both ends. This drive towards least collaborative effort in interaction has been evidenced in different ways: speakers reduce the length of utterances more when they receive positive and concurrent feedback and they use less complex lexicons (and structures) within communicative contexts—i.e., either during dialogue or in isolate production but with the intent to reach an audience different from oneself.

The studies reviewed thus suggest that communicative interaction can have an effect on the complexity of linguistic systems, but there is no systematic comparison between actual communicative interaction and isolate production in the reviewed repeated-reference tasks and in the written modality (for a systematic comparison in the spoken modality, see [Murfit & McAllister 2001](#); for a comparison in the written modality but without repeated usage, see [Traxler & Gernsbacher 1992](#)). In this study I aim to systematically explore the effect of communicative interaction on linguistic complexity in a repeated-reference task, where participants are asked to repeatedly produce descriptions for meanings with the same features. I will compare production in isolation to dyadic communicative interaction (i.e., in two independent experimental conditions). Crucially, the stimuli and the set-up I utilise are the same as the ones implemented at each generation in Experiments 1 and 2 in Chapter 2. This is crucial because I want to test the

validity of the instrumentalisation of the computer-mediated communicative interaction in artificial language studies. I can then relate the results in this experiment to those in Experiments 1 and 2 and shed light on the observed differences between communicative interaction and isolate production in those: communicative interaction facilitates the conventionalisation of transparent and isomorphic form-meaning mappings (i.e., compositional structure) and gives rise to rather simple lexicons early on (what we measure as paradigmatic entropy in section 2.6.2). Only if we know how the operationalisation of communicative interaction in the experimental design influences native speakers' behaviour, can we fully evaluate the effects of the design on non-native performance in artificial language experiments. Comparing communicative interaction to isolate production, the following hypotheses will be tested in the present study:

H1: Greater reduction of description length. If communicative interaction affords higher efficiency in production than mere repetition and/or recency of production, I expect the descriptions to become shorter by trial during the course of communicative interaction than in isolate production (as has been shown in previous studies, e.g. Krauss & Weinheimer 1966).

H2: Lower linguistic complexity. If during communicative interaction interlocutors are driven to maximise efficiency balancing effort in comprehension and production, I expect overall lower linguistic complexity—both lexical and structural—in dyadic communicative interaction than in isolate production (as suggested in previous studies, e.g. Fussell & Krauss 1989; Michel 2011). In order to ease production and comprehension, interlocutors will be faithful to lexical choices which proved to be successful in previous events. Moreover, as interlocutors learn about the amount of information required to successfully convey a given meaning, they can eliminate any linguistic redundancy. Overall, I expect simpler lexicons and structures as well as a higher percentage of content words. Note that unlike H1, H2 makes predictions about the resulting description systems and not about their dynamics during the course of production.

H3: Higher lexical similarity. If interlocutors attune to the audience to ease comprehension effort, I expect more common forms to be used in communicative interaction than in isolate production (as suggested in Fussell & Krauss 1989). I expect to see a higher

overlap across lexicons produced during communicative interaction. As well as H2, H3 makes predictions about entire description systems.

The general contribution of the present study is thus to systematically test the effect of communicative interaction versus isolate production on linguistic complexity at the different levels discussed (i.e., the length of descriptions and linguistic complexity, both lexical and structural) in the written modality. Additionally, this study tests the validity of the simplified computer-mediated model to operationalise communicative interaction in artificial language experiments, in particular the ones presented in Chapter 2.

3.6 Materials and methods

3.6.1 Participants

Thirty native English speakers (aged between 18 and 40, mean age 25.5) were recruited from the University of Edinburgh's Careers Service database of vacancies. Each was paid £7. Nine participants were assigned to the Isolates condition and 18 to the Dyads condition (i.e., nine dyads in total); the data from a further three participants (one pair in Dyads and one participant in Isolates) were excluded from analysis as they did not complete the experiment adequately².

3.6.2 Stimuli

Participants were asked to describe scenes of motion events in English. We used the same set of scenes as in Experiments 1-3 in Chapter 2 (see section 2.2.1.2).

3.6.3 Procedure

Participants were presented with the set of 80 target videos (one at a time, in random order) and were asked to describe them. The experiment was carried out at a computer terminal through a video game interface developed in Python 2.7, which made use of the PsychoPy and Pygame libraries (Peirce 2007, 2009; Pygame Community 2009). Participants sat in individual booths. The experimental sessions lasted approximately 30 and 45 min for the Isolates and the Dyads

²The data from a pair of participants in the Dyads condition were excluded because they used the communicative set-up to send messages unrelated to the experiment's stimuli. The data from one further participant in the Isolate conditions were excluded because they interrupted the experiment before completion.

conditions respectively. All responses were entered using the keyboard. Before participants started the experiment, they were given detailed instructions by the experimenter on how to proceed during the experiment (in both oral and written form).

3.6.4 Experimental conditions

I designed two conditions, one in which participants described the videos one after another on their own (Isolates condition) and another one in which pairs of participants took turns to describe the set of videos to one another and received full feedback on the success of communication after each trial (Dyads condition). Instructions varied accordingly.

Isolates Participants were presented with a different scene at each trial (80 in total) and were asked to provide a description for it (only alphabetic characters and spaces were allowed). As in Chapter 2, I introduced an explicit demand for unique descriptions. Participants were told beforehand that the computer would interrupt them to ask for another description if they had already typed that same exact string to describe a different motion event during the experiment.

Dyads Participants took turns to describe scenes in a dyadic communication task. At each trial (80 in total, one per scene), one participant was assigned the role of sender and the other that of the receiver. The sender typed in a description for a picture, which was sent to the receiver. The receiver saw the description and was asked to select the video that best matched a description out of an array of four (three foils selected randomly plus the target). Both participants received feedback after each trial: they were presented with the target scene, the selected scene and the description (on a green or red background according to their communicative success or failure respectively). The roles of sender and receiver were swapped after each trial. Each participant in a dyad described 40 different videos (half of the total set of 80), balanced to contain instances of all meaning features in roughly the same amounts.

3.6.5 Measures

3.6.5.1 Description length

The reduction of description length through iterated usage is used in interaction studies to measure the increase in efficiency accompanied by processes of conventionalisation during

communication (e.g. H. H. Clark & Wilkes-Gibbs 1986; Krauss & Weinheimer 1964). Description length was calculated in three different ways: by the number of characters, by the number of words (a string of characters between spaces in a sentence) and by the number of phrasal nodes. The number of characters and words partly characterises the syntagmatic complexity of a description, whilst the number of phrasal nodes, partially characterises the hierarchical complexity of a description. Phrasal nodes for each description were obtained automatically using the Berkeley constituent parser (Petrov, Barrett, Thibaux, & Klein 2006; Petrov & Klein 2007) through the Python-Enabled Berkeley Parser (Bengfort 2014). I then manually corrected for wrong class assignments (e.g., if NP was erroneously tagged as a VP). Any other errors were left unmodified and thus constitute a source of noise in the measure, but nevertheless comparable across systems and conditions. It is worth noting that the output does not include any of the following functional phrasal categories: CP, IP³, little vP, or DP. Further phrase types such as AdjPs are not considered either and thus Det Adj N, for instance, only has a unique parent NP (i.e., no DP and AdjP parents).

I measured description length in three different ways because the reduction of one by trial does not necessarily correlate with the reduction of the others. Participants can reduce the number of words leaving the number of phrasal nodes intact if, for instance, they eliminate Det and/or Adj within NPs (e.g., from ‘a pink circle’ to ‘a circle’) as the structural analysis used in this study does not distinguish AdjPs or DPs (see section 3.6.5.2). On the other hand, the description length by character can be reduced leaving the number of words intact. This can happen in at least three different ways: by the reduction of morphological hedges (e.g. “a pinkish square” becomes “a pink square”), by the use of synonymous lexical variants with different word lengths (e.g. “circumnavigate” vs “go around”), and by the clipping of words (e.g. “square” might become “sq”, “sq” or even “s”)—particularly afforded by the written modality.

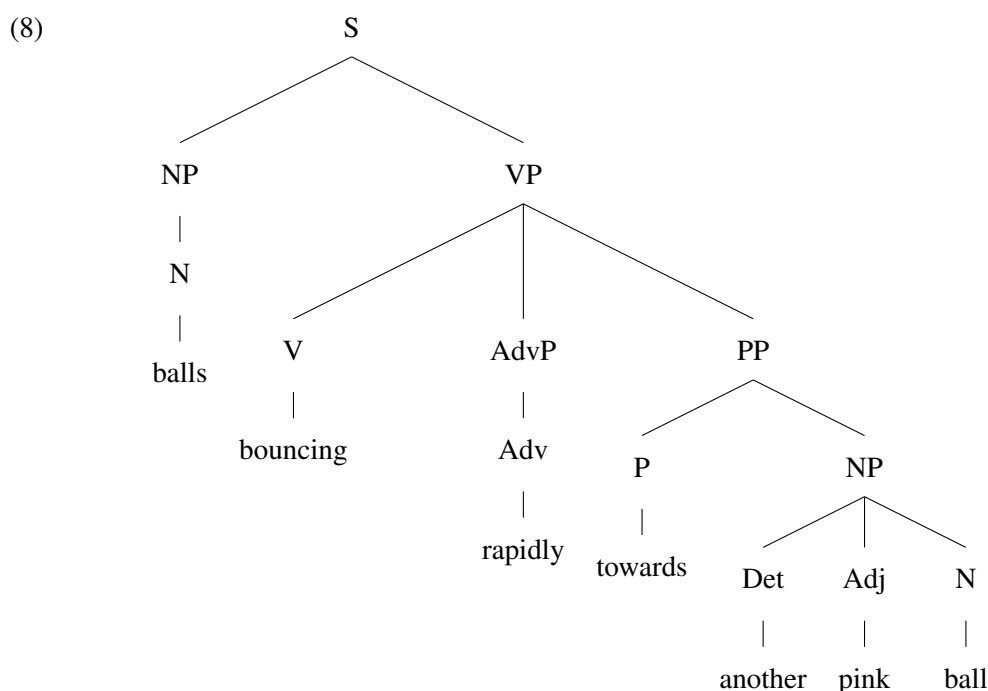
3.6.5.2 Structural complexity

The structural complexity of a given description was measured by means of the number of nodes per description, i.e., non-terminal nodes (phrases) as well as terminal nodes (words). Thus structural complexity of a single description in this case equates to the sum of two of the

³Instead of being a nucleus of an IP, an auxiliary is the nucleus of a VP and a sister of a separate VP.

previously described types of description length, i.e., by word (syntagmatic complexity) and by phrasal node (hierarchical complexity). However, unlike with description length, I will use this measure to compare whole description systems: the structural complexity of a system will be measured as the average structural complexity of all descriptions within a system.

The syntactic tree in (8) illustrates how the Berkeley Parser could analyse the example description “balls bouncing rapidly towards another pink ball”. In this analysis, five phrasal nodes are obtained (excluding the matrix sentence node *S*) for the seven-word description. The structural complexity score for this given description would then be $5 + 7 = 12$.



3.6.5.3 Lexical complexity

For each system of descriptions (i.e., the set of descriptions provided by an isolate participant or a dyad) I extracted its lexicon (i.e., all word tokens comprised across all descriptions). Within the lexicon, I divided words into content and functional categories following the list provided in Cook (1988)⁴. I then separated the content roots from the affixes of content words.

⁴ List of function words (without apostrophes): a, about, above, after, after, again, against, ago, ahead, all, almost, almost, along, already, also, although, always, am, among, an, and, any, are, arent, around, as, at, away, backward, backwards, be, because, before, behind, below, beneath, beside, between, both, but, by, can, cannot, cant, cause, cos, could, couldnt, despite, did, didnt, do, does, doesnt, dont, during, each, either, even, ever, every, except, for, forward, from, had, hadnt, has, hasnt, have, havent, he, her, here, hers, herself, him, himself, his, how, however, I, if, in, inside, inspite, instead, into, is, isnt, it, its, itself, just, least, less, like, many, may, maynt, me, might, mightnt, mine, more, most, much, must, mustnt, my, myself, near, need, neednt, needs, neither, never, no, none, nor, not, now, of, off, often, on, once, only, onto, or, ought, oughtnt, our, ours, ourselves, out, outside, over,

Description	Content words	Function words	Content roots	Affixes
balls bouncing rapidly towards another pink ball	{balls, bouncing, rapidly, pink, ball}	{towards, another}	{ball, bounce, rapid, pink, ball}	{-s, -ing, -ly}

Table 3.1: Example of the partition into the different sets of units used in lexical complexity and similarity measures.

We extracted content roots automatically using PyStemmer (Boulton 2013), a Python interface for the Snowball stemming algorithms (outputs were corrected manually for errors before running any analysis). Affixes were then extracted by subtracting lexical roots from words⁵ and manually corrected. Table 3.1 provides an example of the different sets of units described.

Percentage of content words The percentage of content words in each description system was calculated by dividing the number of content words (i.e., not in the list of function words in footnote 4) by the total number of word tokens.

Lexical richness For each description system I obtained three separate lexical richness scores: one for content roots, one for function words and one for affixes (i.e., inflectional and derivational morphology). Lexical richness was measured by Guiraud’s Index (Guiraud 1954), which is calculated by the number of types divided by the square root of the number of tokens. Guiraud’s Index adjusts type-token ratios to take into account differences in the number of tokens and it has shown to be one of the most robust measures of lexical richness in corpus studies (Hout & Vermeer 2007). For this reason, and also in order to use a measure comparable to those used in natural language dialogue studies (e.g. Michel 2011), I calculate Guiraud’s Index rather than the Shannon entropy used in Chapter 2 to measure the paradigmatic complexity of lexical systems.

past, perhaps, quite, rather, seldom, several, shall, shant, she, should, shouldnt, since, so, some, sometimes, soon, than, that, the, their, theirs, them, themselves, then, there, therefore, these, they, this, those, though, through, thus, till, to, together, too, towards, under, unless, until, up, upon, us, used, usednt, usent, usually, very, was, wasnt, we, well, were, werent, what, when, where, whether, which, while, whilst, who, whom, whose, why, will, with, without, wont, would, wouldnt, yet, you, your, yours, yourself, yourselves.

⁵Note that stem changes are not taken into account in our analysis as inflectional morphology and were counted as different lexical roots.

3.6.5.4 Lexical similarity

I calculated the lexical similarity between the set of lexical units contained in description systems within a given condition. I calculated lexical similarity between types of three different units separately: content roots, function words and affixes. I thus attributed three individual values of similarity per system, one per unit. Each value was defined by the average of all the comparisons between a given system produced by a participant and each of the other systems produced by the remaining participants in the same condition. Lexical similarity between sets of units was calculated using the Jaccard Index (Jaccard 1912). The Jaccard index measures similarity between finite sample sets (i.e., the lexical types), and it is defined as the size of the intersection divided by the size of the union of the sample sets. All output values fall between 0 and 1, the minimum and the maximum similarity scores respectively.

3.6.5.5 Expressivity

I measured the expressivity of each description by the proportion of encoded meaning features which are relevant to the meaning space (see stimuli in 3.6.2). I considered a maximum of six and four features to be encoded in transitive (with two objects) and intransitive (only one object) events respectively⁶. Therefore, for each transitive event, the shape and number of both objects need to be encoded ($2 \times \textit{Shape} + 2 \times \textit{Number} = 4$ features) as well as the type of motion in which the focal objects move (one feature) and the aspect of the motion (one feature). For each intransitive event, only the shape and number of one object as well as motion and aspect features need to be encoded. If the example description in (8) was produced for the scene in Figure 3.1 (i.e., [*FocalCircle*, *FocalPlural*, *AnchorCircle*, *AnchorSingular*, *Bounce*, *Terminated*]), the description would obtain an expressivity score of $5/6 = 0.8\bar{3}$; it describes unambiguously all but one feature, i.e., the terminated aspect of the motion event.

⁶I did not consider Anchoring; it would be difficult to determine the grammatical elements that would encode it in English.

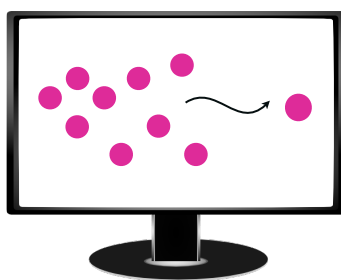


Figure 3.1: Example scene described by “balls bouncing rapidly towards another pink ball”.

3.7 Analyses and results

3.7.1 Description length

I used the *lme4* package (Bates et al. 2015) developed in R (R Core Team 2015) to perform linear mixed-effects analyses to explore the relationship between the length of descriptions and the trial number in which they were produced. P-values were calculated using *lmerTest* (Kuznetsova, Brockhoff, & Christensen 2015). As discussed in section 3.6.5.1, I calculated the length of descriptions by character, by word and by phrasal node. Three different models were run, one for each of these dependent variables, with the same mixed-effects structure. All dependent variables (count data) were log-transformed⁷. As fixed effects I included Trial Number, Condition (two levels: Isolates as reference, and Dyads). As random effects I included intercepts for Description System (one per isolate or dyad) with Interlocutor nested (to control for differences between participants within dyads). I also included a random intercept for Transitivity (i.e., whether meanings included both focal and anchor objects) to control for the effect of the number of meaning features on the length of the descriptions⁸. Figure 3.2 shows the experimental data by condition with the fitted values for the fixed effects overlaid; we observe a general tendency towards the reduction of description length over trials.

In the analysis of the relationship between description length by character and Trial Number (see Figure 3.2, top) I found a significant effect of Trial Number ($\beta = -0.003, SE = 0.0005, p < 0.001$), suggesting that the number of characters per description goes down significantly by trial. There was no significant effect of Condition ($\beta = -0.007, SE = 0.193, p =$

⁷I report the results of a linear model with log-transformed data because negative binomial models resulted in converge warnings and the resulting estimates were equally good ($\chi^2(0) = 0, p = 1$).

⁸I include Transitivity as a random effect because I am not interested in the specific contribution of transitive or intransitive events as much as I am interested in the variation due to Transitivity and in assessing the potency of Description System after accounting for this variation.

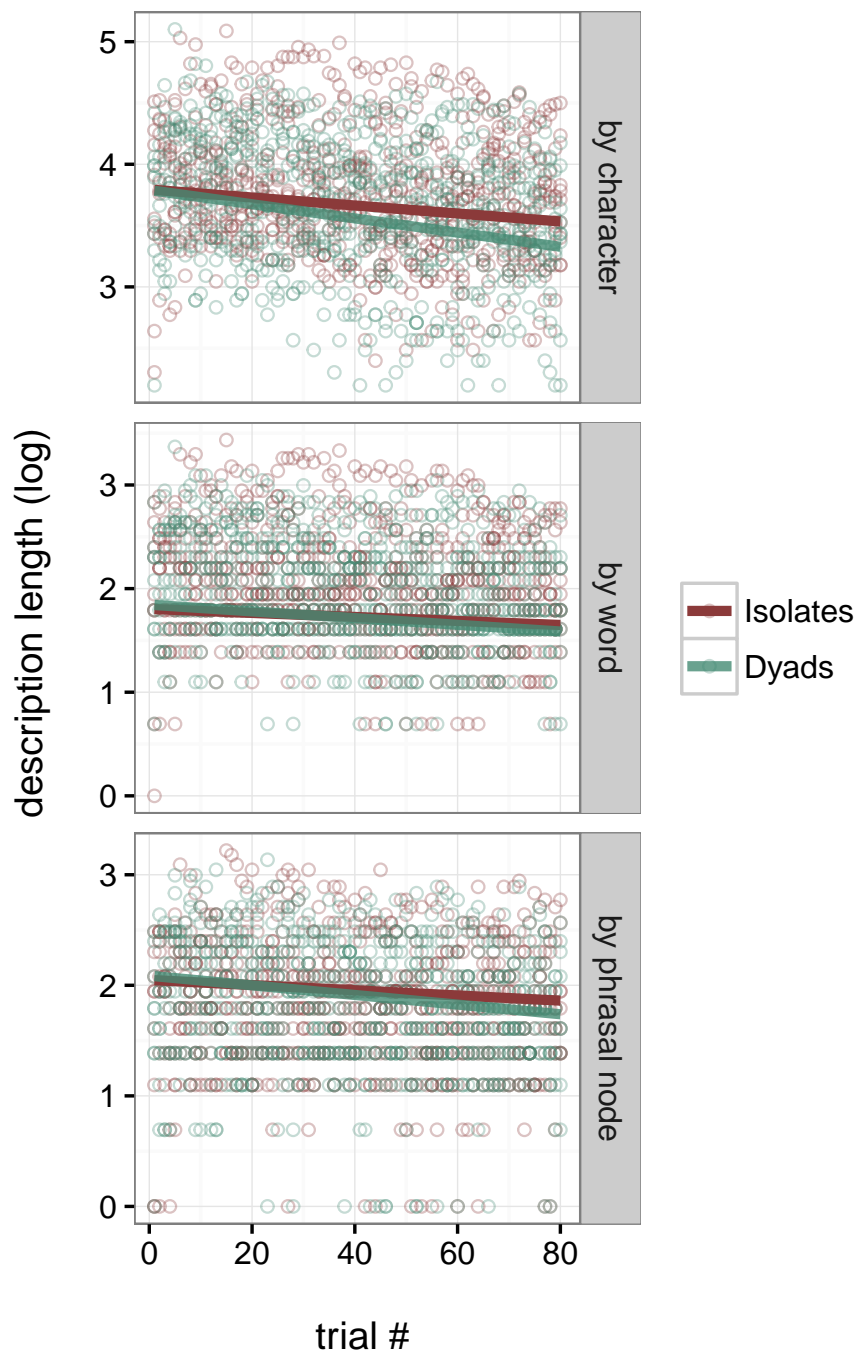


Figure 3.2: Log description lengths by character, word and phrasal node. Lines represent fitted values for the fixed effects estimated by the linear mixed-effects analysis.

0.971), but I found a significant interaction between Trial Number and Condition ($\beta = -0.0025$, $SE = 0.0006$, $p < 0.001$). These results suggest that although Dyads did not produce more characters per description at the first trials, the decrease in length by trial is significantly greater in Dyads than in Isolates. The overall model fit was $R^2_{\text{marginal}} = 0.049$ and $R^2_{\text{conditional}} = 0.734$.

Results from the model with description length by word as the dependent variable (see Figure 3.2, middle) show a significant effect of Trial Number ($\beta = -0.0023$, $SE = 0.0005$, $p < 0.001$), suggesting that the number of words per description decreases significantly by trial. I did not find a significant main effect of Condition ($\beta = 0.0036$, $SE = 0.197$, $p = 0.859$), but there was a significant interaction between Trial Number and Condition ($\beta = -0.002$, $SE = 0.0006$, $p = 0.002$), suggesting that, although description length by word is comparable at the first trial, the decrease in length by trial is significantly greater in Dyads than in Isolates. The overall model fit was $R^2_{\text{marginal}} = 0.024$ and $R^2_{\text{conditional}} = 0.723$.

For description length by phrasal node (see Figure 3.2, bottom) I also found a significant, although weaker, effect of Trial Number ($\beta = -0.0019$, $SE = 0.0006$, $p = 0.001$), suggesting that the number of phrasal nodes decreases significantly by trial. There were no significant effects of Condition ($\beta = 0.0395$, $SE = 0.197$, $p = 0.843$) or its interaction with Trial Number ($\beta = -0.0013$, $SE = 0.0008$, $p = 0.128$), suggesting that description lengths by phrasal node at first trial are comparable across conditions, and so are the slopes by Trial Number. The overall model fit was $R^2_{\text{marginal}} = 0.010$ and $R^2_{\text{conditional}} = 0.616$.

3.7.2 Structural complexity

In order to explore the effect of condition on structural complexity (calculated as described in section 3.6.5.2), I ran a linear mixed-effects model. The dependent variable, i.e., structural complexity (count data), was log-transformed before analysis⁹. As fixed effects, I included Condition alone; and as random effects, intercepts for Description System. The overall model fit was $R^2_{\text{marginal}} = 0.0006$ and $R^2_{\text{conditional}} = 0.594$. I found no effect of Condition on structural complexity ($\beta = -0.028$, $SE = 0.199$, $p = 0.88$).

⁹I report the results from a linear mixed-effects model instead of a negative binomial mixed-effects model for consistency with the analyses carried out in the following section 3.7.1. Additionally, the linear model constituted a better fit ($\chi^2(0) = 7514.7$, $p < 0.001$ and, in both cases, I found no significant effects.

3.7.3 Lexical complexity

Percentage of content words I performed a logistic mixed-effects regression model to explore the effect of condition on the percentage of content words. I included Condition (two levels: Isolates as reference, and Dyads) as the only fixed effect and a sole random intercept for Description System (i.e., each of the systems produced by an isolate participant or a dyad). The overall model fit was $R^2_{\text{marginal}} = 0.0003$ and $R^2_{\text{conditional}} = 0.043$. The model intercept suggests that the percentage of content words in the Isolates condition was high ($\beta = 0.78, SE = 0.131, p < 0.001$), and no significant difference between conditions was found ($\beta = 0.066, SE = 0.185, p = 0.72$). Results thus suggest that the percentage of content words was comparably high across conditions.

Lexical richness Lexical richness scores (calculated as described in 3.6.5.3) are shown in Figure 3.3. I ran a linear mixed-effects model to explore the effect of condition on lexical richness. As fixed effects I included Condition (two levels: Isolates as reference, and Dyads) and Lexical Unit (three levels: Content Roots as reference, Function Words and Affixes) with an interaction term. As random effects, I added intercepts for Description System. The overall model fit was $R^2_{\text{marginal}} = 0.514$ and $R^2_{\text{conditional}} = 0.752$. Results show that lexical richness is higher for content roots than for function words ($\beta = -0.780, SE = 0.193, p < 0.001$) or affixes ($\beta = -1.615, SE = 0.193, p < 0.001$), suggesting that, as expected, function words and affixes are more productive than content roots within lexicons. Most importantly, I found a significant effect of Condition ($\beta = -1.012, SE = 0.271, p < 0.001$), suggesting that lexical richness of content roots is higher in the Isolates than in the Dyads condition. The same effect of Condition was maintained for the sets of function words ($\beta = 0.431, SE = 0.274, p = 0.125$) but not for affixes ($\beta = 0.887, SE = 0.274, p = 0.003$). These results suggest that lexical richness is higher in Isolates for the sets of content roots and function words but not for affixes, where it is comparable to that in Dyads.

3.7.4 Lexical similarity

Lexical similarity scores (calculated as described in 3.6.5.4) are shown in Figure 3.3. In order to avoid collinearity, I ran three separate linear models to explore the effect of condition on lexical similarity: one model to explore this effect within sets of content-root types, one for the sets of

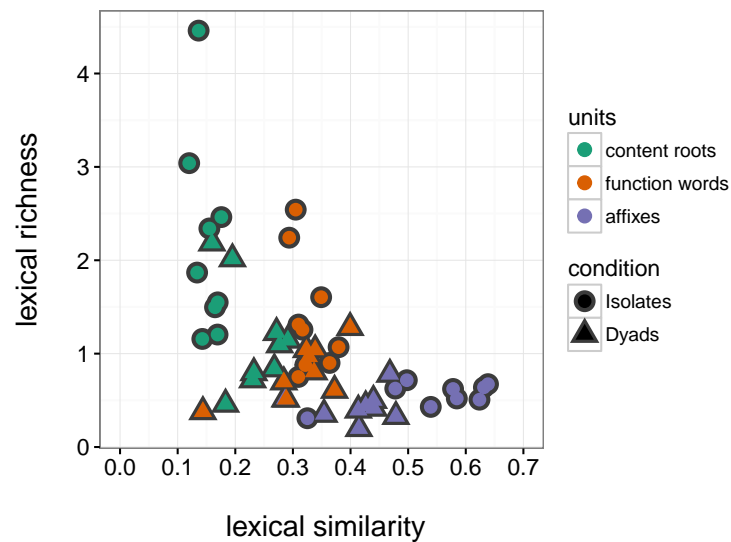


Figure 3.3: Relationship between lexical richness (y-axis) and lexical similarity (x-axis) for each description system in the Isolates (circles) and Dyads (triangles) conditions. The colours green, red and purple correspond to the measures for sets of content roots, function words and affixes respectively. Focusing on the y-axis, there is a general tendency for richer lexical sets in the Isolates condition. Looking at the relationship between lexical richness and similarity, we observe they are generally negatively correlated: i.e., the richer a lexical set is, the less similar it is to the other sets within the same condition.

function-word types and one for the sets of affix types. Each model shared the same structure: they contained Condition and Lexical Richness as predictors. I included Lexical Richness as a predictor because, as observed in Figure 3.3, I expect lexical richness and lexical similarity to be (negatively) correlated and I want to tease apart the effects of condition on lexical similarity that are not simply explained by differences in lexical richness. The model fits were $R^2 = 0.591$, $R^2 = 0.375$ and $R^2 = 0.451$ for the sets of content roots, function words and affixes respectively (all R^2 are adjusted). Results for content-root types suggest that their similarity within condition is significantly higher for Dyads ($\beta = 0.099, SE = 0.039, p = 0.023$), an effect that does not change in interaction with Lexical Richness ($\beta = -0.022, SE = 0.0242, p = 0.374$). Results for function-word types suggest that their similarity within condition is significantly lower for Dyads ($\beta = -0.191, SE = 0.058, p = 0.005$); however, they also show a significant interaction ($\beta = 0.201, SE = 0.057, p = 0.003$), suggesting that the difference between conditions can be predicted from its interaction with lexical richness. Results for affix types showed no significant effects: similarity between sets of affixational morphology within condition is comparable between conditions ($\beta = 0.092, SE = 0.144, p = 0.533$). Altogether these results suggest that only the within condition similarity of content roots is higher for Dyads than for Isolates, thus

hinting that more common content roots were used in dyadic communicative interaction.

3.7.5 Expressivity

We observed that description systems in Isolates are lexically richer and that description lengths are not as reduced as in Dyads with iterated production. However, these differences between conditions could be driven by differences in expressivity (calculated as described in section 3.6.5.5); participants in the Isolates condition, unlike those in Dyads, are not allowed to reuse descriptions. In order to rule out this possibility, I ran a logistic mixed-effects regression model to test the effect of condition on expressivity. I included a sole fixed effect for Condition and a random effect for Description System. The overall model fit was $R^2_{\text{marginal}} = 0.0003$ and $R^2_{\text{conditional}} = 0.043$. The model intercept suggests that expressivity is high in the Isolates conditions ($\beta = 1.471, SE = 0.134$), i.e., the average proportion of expressivity in description systems is $\hat{p} = 0.813$. No effect of Condition was found ($\beta = 0.066, SE = 0.189, p = 0.729$), suggesting that expressivity is comparably high across conditions. These results thus suggest that differences in expressivity cannot explain the differences we find in lexical richness and description length between conditions.

It is worth mentioning that participants in the Isolates condition did not generally try to reuse descriptions for different meanings, only one third of participants attempted to reuse a description once (and then were asked to introduce an alternative description by the computer). Participants in the Dyads condition—who in contrast could reuse descriptions—also tended to produce unique descriptions for each scene; out of a total of 80 different scenes, the median number of unique descriptions per system was 79 (Median Absolute Deviation (MAD)= 0.012); only two dyads were below the median, with 67 and 78 unique descriptions, and four above with 80. Given the high expressivity scores, it is no surprise that communicative accuracy (i.e., the proportion of successful communication trials) in the Dyads condition was at ceiling (M=0.953, SD=0.036).

Altogether, these results suggest that the differences across conditions are not driven (at least uniquely) by the constraint to produce unique descriptions in the Isolates condition. However, asking participants to produce unique descriptions for a given meaning might have biased them against reusing linguistic units (e.g., words); this bias could be partially driving the differences in lexical richness across conditions. The presence of this constraint in production

against homonymous descriptions was required in order to contrast these findings with the results of the experiments carried out in Chapter 2. Further work—an Isolates condition in which no constraint in production is implemented—is thus required to directly test the effect of this constraint, and to what extent it can predict the differences across conditions.

3.8 Discussion

In this study I tested the effect of communicative interaction on linguistic complexity in a repeated-reference task. It has been proposed that communicative context allows interlocutors to tailor expressions to the needs of the interlocutors during interaction (H. H. Clark & Wilkes-Gibbs 1986; Fusaroli & Tylén 2012; Lewis 1968). This process thus allows speakers to test the minimal effort required to achieve effective communication, ultimately leading to the establishment of efficient communicative systems (H. H. Clark & Wilkes-Gibbs 1986). This drive towards efficiency prompted by the communicative context has been shown to have an effect on the overall linguistic complexity of communication systems (Brennan & Clark 1996; Kanwal et al. 2017). However, there was no systematic comparison between the differences in efficiency between isolate production and communicative interaction in a repeated-reference tasks and in the written modality (cf. Michel 2011; Murfitt & McAllister 2001). This study was designed to provide this systematic comparison. Specifically, I set out to test three hypotheses (see beginning of section 3.5): in comparison to production in isolation I expected communicative interaction to lead to greater reduction of description length, greater lexical similarity within conditions and lower overall linguistic complexity. In the following pages I discuss the findings for each of these hypotheses.

3.8.1 The effect of communicative interaction on efficiency in production: description length over repeated use

Descriptions shortened with repeated use both in isolate production and in dyadic communicative interaction, suggesting that frequency and/or recency of production trigger a drive to reduce effort. Over repeated use, the numbers of phrasal nodes, words and character were significantly cut down in descriptions, suggesting that descriptions' structural and lexical complexity were both reduced by trial. It is worth noting that we observe the reduction of description length

even though each description matches a unique scene, suggesting that the repetition of the features of a scene is enough to trigger similar effects to those observed in previous studies studying the development of referring expressions (H. H. Clark & Wilkes-Gibbs 1986; Krauss & Weinheimer 1966), where participants repeatedly provide descriptions for the same object in isolation. Additionally, this study follows previous work in demonstrating that the descriptions of novel referents become more efficient with repeated use in dyads (Krauss & Weinheimer 1964, 1966). Dyadic communicative interaction leads to a significantly higher reduction of description length than isolate production, suggesting that the communicative context licenses efficiency further; differences between conditions cannot be due to repeated use and/or recency of production alone. However, these differences in communicative efficiency between conditions are only observed in the reduction of words and characters; the contraction of phrasal nodes by trial is comparable in isolates and dyads. This might be due to isolates using more modifiers within NPs, e.g., adjectives, which are not reflected by the number of phrasal nodes in the Berkeley parser used. The average number of adjectives in the systems produced by isolates and dyads were approximately 85 and 53 respectively; although due to the great variability of the data and the small sample size a negative binomial regression model shows no significant difference between conditions ($\beta = -0.480, SE = 0.404, p = 0.235$), these average differences could be influencing the overall non-significant difference between conditions of the reduction of phrasal nodes by trial. If this were the case, it could be explained by the expectancy of higher reduction of redundant information in dyadic communicative interaction.

This is the first study which shows that shortening of descriptions results both from isolate production and communicative interaction and that it is significantly higher in communicative interaction (cf. Fussell & Krauss 1989; Krauss & Weinheimer 1966). Unlike in Krauss and Weinheimer (1966), where participants do not increase efficiency in production in the absence of positive or concurrent feedback because they receive negative feedback 50% of the time, participants in this study's Isolate condition do not receive any feedback whatsoever and thus they are not questioned on the adequacy of their descriptions. Consequently, regardless of the lack of positive or concurrent feedback, participants are not held back from increasing efficiency in production. Unlike in Fussell and Krauss (1989), where speakers produce longer descriptions when they think their utterances are intended for another person, this study's social condition includes actual communicative interaction. Although this study never includes concurrent

feedback (i.e., participants only receive feedback after sending a message), participants take turns to describe scenes to each other. At each trial, the addressee only has access to the final description, but eventually both interlocutors are provided with full feedback on their intended and matched scenes. This type of feedback provides much of the information about the interlocutor's intention and knowledge disclosed during concurrent feedback. Moreover, the effects of intended audience are moulded by interaction as interlocutors update their knowledge about the audience and the audience's knowledge. Without the presence of an interlocutor, speakers might try to provide as much information as they might feel necessary but only the interaction with interlocutors will allow the speaker to know how much information is in fact required. This would explain the differences regarding description length I observed in my results and the ones in [Fussell and Krauss \(1989\)](#).

Neither the shortening of descriptions nor the differences between conditions can be due to time constraints. There are no time restrictions; speakers can take as long as they want and rewrite sentences before the submission of the final description. Participants, nevertheless, could be producing shorter sentences because they want to finish the experiment quickly, but it would not explain differences between conditions. The results show that, where there is communication, communicative accuracy is at ceiling throughout the experiment and no difference can be found in expressivity between Isolates and Dyads, which suggests that participants are developing more efficient descriptions rather than just producing descriptions that are shorter at the expense of not encoding enough information.

The difference observed between conditions supports a collaborative model in the development of communication systems where the build-up of shared knowledge through interaction is crucial ([Brennan & Clark 1996](#); [H. H. Clark & Marshall 1981](#); [Fusaroli & Tylén 2012](#); [Garrod & Doherty 1994](#); [Wilkes-Gibbs & Clark 1992](#)). Only during interaction, with the provision of feedback, do participants update their shared knowledge, which allows them to quickly develop more efficient systems reliably.

3.8.2 The effect of communicative interaction on efficiency in comprehension: lexical similarity

This study provides behavioural support for the well-established model of communication in which shared knowledge is exploited to ease production as well as comprehension. Dyads

adapted their descriptions to their addressees inasmuch as they opted for common content roots. We observe this in the present study from the higher lexical similarity obtained between systems within the Dyads condition in comparison to those in Isolates—which is not merely driven by differences in lexical richness. These results suggest that interlocutors' lexical choices were biased by the communicative context; they chose more transparent content roots to ease comprehension and assure communicative success (Fussell & Krauss 1989). Without a history of interaction interlocutors initiate communicative events in the Dyads condition without shared knowledge specific to the situation but with linguistic as well as general knowledge; participants who initiate a communicative event make use of this knowledge to choose content words they conceive to be shared amongst English speakers in their community and therefore more likely to be understood by their interlocutors. Participants in the Isolates condition, on the other hand, do not get such an advantage from tailoring their lexical choice to their general knowledge.

3.8.3 The effect of communicative interaction on linguistic complexity

I found partial support for the hypothesised lower linguistic complexity in communicative interaction in comparison to isolate production; systems produced in dyads rather than in isolation contain lower lexical complexity. However, structural complexity was comparable between conditions. This study follows previous work in demonstrating that lexicons produced in isolation were richer than those in dyadic communicative interaction, i.e., they were more diverse and contained more lexical units of lower frequency (Fussell & Krauss 1989; Krauss, Vivekananthan, & Weinheimer 1968; Michel 2011). The difference in lexical richness was provided by content roots and function words but not by affixational morphology (inflectional and derivational), i.e., lexicons contain comparably rich affixational morphology across conditions but Isolates produce richer sets of content roots and function words. Although the sets of function words are richer in Isolates than in Dyads, the percentage of function words in relation to the overall tokens produced was comparable across conditions. These results suggest that function and content words were produced in similar ratios across conditions but function as well as content words (at least content roots) were more diverse in Isolates than in Dyads. However, this difference between conditions could be partially driven by a bias to not repeat lexical units in the Isolates condition given the constraint to produce unique descriptions. Therefore,

although our analysis suggest that this constraint is unlikely to be the sole driver of the differences observed, further work needs to be done to directly test the effect of the implemented constraint on lexical richness.

On the other hand, our results do not follow Michel (2011) in showing lower structural complexity in dyadic communicative interaction. Unlike Michel (2011), this study analyses production in the written modality and in a repeated-reference task through a computer interface, where participants' responses are more restricted. Differences between lexical and structural complexity in our study suggest that the drive towards least collaborative effort biases participants to reduce the variability of the lexicon and thus to stick to one-to-one form-meaning mappings negotiated during interaction but not necessarily to produce structurally simpler descriptions. Moreover, results show that higher reduction of words by trial does not necessarily lead to overall significant differences in structural complexity across conditions—probably also an artefact of the lack of intra-NP phrases in the analysis.

3.9 Conclusion

This study follows previous studies in demonstrating that the inclusion of communicative interaction, where feedback is provided, leads to more efficient communicative systems with lower overall linguistic complexity, in particular, shorter descriptions and simpler lexicons (H. H. Clark 1996; H. H. Clark & Wilkes-Gibbs 1986; Krauss & Weinheimer 1966; Michel 2011). These results further support the idea that communicative interaction leads to participants engaging in a process of *least collaborative effort*, in which they exploit shared knowledge to minimise the total effort spent during interaction—in both production and comprehension (H. H. Clark & Wilkes-Gibbs 1986; Davies 2006). I have shown that in comparison to isolate production, communicative interaction leads to increased efficiency or least effort in production; i.e. dyads contract the number of words and characters in descriptions by trial significantly more than isolates. The communicative context drives interlocutors to reduce the effort in comprehension as well as in production; lexicons are simpler (i.e., they contain fewer lexical units and those units are used more frequently) and contain common content words (i.e., content roots are more similar across dyads than they are across isolates).

Altogether, these results support an understanding of communicative interaction as a joint

activity whereby participants use feedback to align on a system of shared knowledge which they can exploit to increase communicative efficiency. Within the mechanisms whereby interlocutors increase efficiency, our results suggest that there are more than low-level priming mechanism at play. Interlocutors seem to attune their behaviour to the communicative context in ways which cannot be straightforwardly explained by competences which are not higher-level—as Pickering and Garrod (2004) would describe mechanisms within theory of mind; the higher lexical similarity obtained from systems produced in communicative contexts cannot be explained by priming alone. Generally, although the current study does not aim to tease apart implicit from explicit mechanisms, results suggest that for an automatic account of alignment to provide satisfactory explanations for the differences between conditions found in this study, it should provide a stronger model that predicts similarities as well as differences between priming in isolate production and reciprocal priming (cf. Pickering & Garrod 2004).

This study further supports the ecological validity of the simplified model of communicative interaction used in Chapter 2 and further supports the implied effect of coordination on facilitating the evolution of linguistic structure, in combination with expressivity and learnability pressures. It is through coordination that linguistic conventions are established; moreover, the drive to reduce collaborative effort leads to simpler systems being used, which increases learnability and accuracy and thus allows for higher systematisation.

Chapter 4

The effect of level-specific linguistic variation on regularisation behaviour

4.1 Introduction

Variation is ubiquitous in language; it can be found at all linguistic levels, from phonology (allophones), to the lexicon (synonyms), to morphology (allomorphs) and syntax (paraphrases) (Labov 1972). However, linguistic variation *tends* to be (at least partially) conditioned: the choice of variant is predictable by some aspect of the social or linguistic context. For instance, English exhibits two different realisations of comparative and superlative adjectives: a synthetic one via the use of the suffixes *-er* and *-est* respectively (e.g., ‘faster/fastest’), and an analytic one via the independent forms *more* and *most* preceding the adjective (e.g., ‘more/most intelligent’). However, this type of morphological variation is mostly deterministic: monosyllabic and bisyllabic adjectives are more likely to be derived synthetically than longer adjectives, which are derived analytically. Other types of variation are more probabilistic in nature: in some situations speakers are more likely to produce certain variants than in others. For instance, English exhibits probabilistic variation in progressing tense allomorphy, e.g., between [ɪn] or [ɪŋ], and speakers’ choice of variant varies according to register or their social status (Labov 1972; Shuy, Wolfram, & Riley 1967; Trudgill 1974). Thus even in these probabilistic cases, variation is still conditioned and thus somewhat predictable.

Despite the strong tendency for variation to be conditioned, there are some circumstances in which *unconditioned* variation is nevertheless found. For instance, unconditioned variation

can be found when new variants are first introduced into an established system, or when input is limited and conventions are still not established, as in contexts of language formation (Good 2012; Kouwenberg & Singler 2009; Siegel 1997; Velupillai 2015). Under these circumstances, there is substantial evidence that learners tend to reduce or remove such variation, i.e., they regularise the system (Newport 1999; Siegel 1997). This can be achieved either by reducing or removing competing variants or conditioning variant choice on the context (Ferdinand, Kirby, & Smith 2017). Furthermore, regularisation can be seen both in individuals and over time as languages change or become more conventionalised. (Hare & Elman 1995; Kouwenberg & Singler 2009).

Indeed, regularisation (i.e., the reduction, elimination or conditioning of variation) has been documented extensively in natural languages across linguistic levels including phonology, morphology, syntax and the lexicon; i.e., in language acquisition (Fraser et al. 1963; Newport 1999; Ross & Newport 1996; Singleton & Newport 2004), language change (Hare & Elman 1995; Schilling-Estes & Wolfram 1994; van Trijp 2013), and in language formation (Bickerton 2015; DeGraff 1999; McWhorter 2005; Senghas & Coppola 2001; Senghas, Newport, & Supalla 1997; Siegel 2004; Spears 2008; Winford 2003; Yule 1996). In addition, experimental studies involving artificial language learning techniques report regularisation behaviour during learning and production of probabilistic variation of diverse linguistic elements, across different linguistic levels (Culbertson et al. 2012; Fehér, Ritt, Wonnacott, & Smith 2016; Ferdinand et al. 2017; Hudson Kam 2015; Hudson Kam & Chang 2009; Hudson Kam & Newport 2001, 2005, 2009; Perfors 2012a, 2012b; Perfors & Burns 2010; Real & Griffiths 2009; Schumacher, Pierrehumbert, & LaShell 2014; K. Smith & Wonnacott 2010; Wonnacott & Newport 2005). These studies provide evidence for how biases in language learning and use might interact to shape the kind of variation found in natural language.

In a now classic study, Hudson Kam and Newport (2005) used artificial language learning to investigate developmental differences in regularisation behaviour. The artificial language taught to participants contained unconditioned variation in the presence or absence of determiners. While adult learners were found to roughly match the level of probabilistic variation in the input, children's productions were more regular and governed by idiosyncratic but invariant patterns. Using the same paradigm, Hudson Kam and Newport (2009) tested the effect of system complexity on regularisation behaviour. Across a number of conditions, a "main"

determiner was used 60% of the time, but what appeared the remainder of the time was manipulated across conditions; either no determiner occurred, or a number of “noise” determiners were used (either 2, 4, 8 or 16 noise determiners). Differences in adults and children were again found—children were more likely to produce completely deterministic patterns of determiner usage than adults. However, it is important to note that adults’ regularisation behaviour was affected by the complexity of the system. The more noise determiners were present, the more adults regularised the main determiner. These results suggest that adult learners are more likely to regularise complex systems of probabilistic variation.

Building on these earlier studies, a number of other researchers have investigated regularisation behaviour in adults, showing that weak biases for regularisation can be amplified over generations of learners in an iterated learning paradigm (Reali & Griffiths 2009; K. Smith & Wonnacott 2010) and that generalisation can increase regularisation (Wonnacott & Newport 2005). Further, recent studies have suggested that regularisation can also be modulated by the nature of specific structures to be learned. For example, Culbertson et al. (2012) and Culbertson and Newport (2015) show that learners regularise harmonic word order patterns (i.e., with either consistently head-initial or head-final) in the noun phrase, more than non-harmonic patterns. Taken together, these experiments suggest a link between biases active during learning and the scarcity of unconditioned variation in language. Under this view, regularisation at the population level is taken to reflect biases in language learning and/or use at the individual level. While the precise nature of the bias is not yet known, one possibility is that it is driven by a bias for learnability. Regularisation leads to the reduction of variation by favouring one variant over another or by conditioning variants on any aspect of their social or linguistic context; the resulting language is thus simpler and/or more predictable (Culbertson & Kirby 2016). This is in line with findings suggesting that regularisation is modulated by the type of input learners receive. A learner’s response to linguistic variation depends on the nature of the structures in question (Culbertson & Newport 2015; Culbertson et al. 2012) and the complexity of the input learners receive (Hudson Kam & Newport 2009).

4.1.1 Level-specific effects on regularisation

Despite the aforementioned extensive literature investigating regularisation, relatively little is known about whether and how regularisation might *differ* across linguistic levels and units.

While laboratory experiments on regularisation have looked at lexical, morphological and syntactic variation, no study has directly compared them. These studies implicitly assume uniform mechanisms and processes behind regularisation across linguistic levels. However, research on natural language learning and formation suggests the possibility that morphological and syntactic variation may not be treated the same by learners (Good 2015; Siegel 2006; Slobin 1986).

Research on L1 acquisition suggests that the mastery of word order precedes that of morphological inflection: children produce more morphological than word order errors (Bichakjian 1988; Slobin 1966), including over-regularisation (Marcus et al. 1992). Studies looking specifically at the acquisition of variation *in the input* indicate that variation in morphology in children's production is less likely to be target-like and/or is more prone to regularisation (K. L. Miller & Schmitt 2012; Raymond, Healy, McDonnell, & Healy 2009) than variation in syntactic features such as word order (Anderssen, Bentzen, & Westergaard 2010). Similar differences between morphology and word order features seem to exist in L2 acquisition: errors on inflectional morphology are maintained for longer in a learner's interlanguage than those in word order, which are often caused by L1 transfer initially and are soon corrected (Dietrich, Klein, & Noyau 1995; Montrul 2004; Siegel 2006). Furthermore, differences between levels are not only raised in acquisition but also in language attrition; in a study looking at patterns of attrition in Swedish grammar, (Håkansson 1995) observed that whereas expatriate bilingual Swedes produce native-like variation in word order patterns (i.e., without V2 errors), they failed to reproduce NP-internal inflectional morphology (i.e., determiner and adjective inflection for gender, number and definitiveness).

Pidgin/creole studies suggest differences in morphological complexity between pidgins and source languages (McWhorter 2001; Siegel 2004) and on the other hand, they also suggest that pidgins show the word orders we find in languages around the world in roughly comparable proportions (Bakker 2008; Good 2015). Compared to source languages, it is common for early pidgins to initially show poorer inflectional and derivational morphology; however, several studies indicate that word orders tend to be comparably variable (Drechsel 1981; J. D. Nichols 1995; Stefnsson 1909; Thomason 1980; van der Voort 2013). Eskimo pidgins, for example, show a rather variable word order (van der Voort 2013); e.g., although SOV (i.e., the default order in the lexifier West Greenlandic) is prevalent, we find instances of SVO (the default or-

der in the other source languages, i.e., English, Danish, French and Russian) and VSO orders. Similarly, Eskimo pidgins also use both adjective-noun and noun-adjective orders, a mixture of those found in their source languages (van der Voort 1997). Conversely, we find that NP-internal inflectional morphology in source languages does not equally prevail: e.g., Eskimo pidgins omit the possessor inflection found in West Greenlandic (and often in other source languages, e.g., English), leaving the possessor and the thing possessed in bare juxtaposition (van der Voort 1997, 2013). Provided that simplification processes in languages formation—as well as in language change more generally (Bichakjian 1988; Slobin 1986),—reflect individual biases during learning (Lefebvre, White, & Jourdan 2006; Siegel 2006) and/or use (McWhorter 2001; Parkvall 2008), the early loss of inflectional morphology and maintenance of word order in pidginisation would further support differences between linguistic levels in language learning and/or use.

While this evidence is suggestive, it remains an open question whether regularisation applies with uniform strength across linguistic levels or not, and to what extent this is driven by level-specific biases in language learning and use. There are alternative explanations for the suggested differences between regularisation across levels which do not necessarily require global level-specific biases. Firstly, we cannot assume a comparable initial complexity between linguistic paradigms at different levels in input languages. There is no corpus study that tests this but if evidence was found for the general intuition that morphological systems of variation tend to be more complex than word order, resulting differences in simplification processes could then be attributed to these differences in the inherent complexity of the systems; as previously mentioned, there is experimental evidence demonstrating a positive correlation between the complexity of variation systems and regularisation behaviour (Hudson Kam & Chang 2009; Hudson Kam & Newport 2009). Secondly, differences between linguistic levels might be driven by specific traits of a given level in the input languages; language acquisition and language formation is sensitive to the variants in competition in the set of features available to the user (Mufwene 2008; Siegel 2006). Early on in L2 acquisition, for example, language learners tend to favour variants that match their prior linguistic knowledge more closely. Consequently, the amount of overlap between features across source languages could also determine the preservation of a given grammatical feature during pidgin formation (Ansaldo, Matthews, & Lim 2007; McWhorter 2005). If morphological paradigms are more complex than word or-

der, the overlap between full morphological paradigms is less likely than the overlap between word order variation. It is also less probable that all variants in more complex systems are present in the data available to the learner or that they are all salient enough to be reproduced (Mufwene 2008).

In the present study I systematically compare the strength of regularisation across linguistic levels—specifically, morphology and word order—to test the extent to which level-specific biases interact with regularisation behaviour during language learning and use. Because previous literature suggests that system complexity impacts regularisation, I carefully control the complexity of the systems learned (i.e., in terms of input variation). Results demonstrate that when input languages have comparable initial complexities and there are no specific biases targeting variants available only at one level, regularisation behaviour is also comparable across morphology and word order. This suggests a single regularising mechanism at work, with apparent differences among levels likely due to differences in inherent complexity.

4.2 The present study: regularisation behaviour across linguistic levels

In Experiments 5 and 6, I explore regularisation behaviour in individual participants across two linguistic levels: morphology and syntax (word order). I use the methodology developed in Culbertson et al. (2012) (following Hudson Kam & Newport 2005, 2009). Adult learners are exposed to a miniature artificial language featuring an inconsistent mixture of synonymous variants, whose use is determined probabilistically, with no conditioning factors. I assess how learners restructure these input languages in their productions and to what extent that restructuring is comparable when the variation involves morphology and word order. Since I know from previous work that the amount of regularisation can depend on the complexity of the system (particularly for adult learners), I construct input languages which differ *only* in whether variation is in word order or in morphology. The languages are comparable in their inherent complexity, i.e., they are equally variable. In Experiment 7, I include communicative interaction during production using the methodology developed in Fehér, Wonnacott, and Smith (2016) in order to investigate the effect of communicative interaction on the strength of regularisation across linguistic levels.

4.3 Experiment 5: regularisation behaviour across linguistic levels in individuals

4.3.1 Materials and Methods

4.3.1.1 Participants

Fifty-six native English speakers (19 male, 37 female; aged between 18 and 41, mean 23.2) were recruited from the University of Edinburgh's Careers Services database. Participants were paid £6. Twenty-six participants were assigned to the Morphology condition and 26 to the Word Order condition. The data from a further four participants was excluded on the basis of failure to adequately learn the language. This was determined based on a pre-set limit on incorrect responses in the production testing phase.¹

4.3.1.2 Input languages

The input languages were used to describe simple pictures featuring one of two objects—each object consistently mapped to a specific noun, N_1 ('jelpa') and N_2 ('mokte'). Objects appeared either singly or in a pair and could appear either in grey-scale or coloured in blue. All eight pictures are shown in Figure 4.1 (these are a subset of images used in [Culbertson et al. 2012](#)). When the objects appeared singly and in grey-scale, they were considered bare (un-modified) objects, and were accordingly described by a bare Noun (N_1 or N_2). The rest of the descriptions corresponding to objects in pairs and/or in colour were NPs composed of a Noun plus a Num(eral) and/or Adj(ective) modifier, which were presented orthographically and aurally to participants during training. Input languages were small so as to minimise the effort of learning their lexicon and maximise participant's attention to variation.

All lexical items were 5 graphemes/phonemes long and had a neighbourhood density of 0 in the English lexicon. Nouns and modifiers were bisyllabic CVCCV words, but differed slightly in their syllabic structure; the internal cluster for nouns (i.e. 'mokte', 'jelpa') required CVC.CV syllabification while the internal cluster for modifiers was a legal onset (i.e. 'nefri', 'nezno', 'kogla', 'kospu') and thus could be syllabified as CV.CCV. The Levenshtein edit distance (LD)

¹Similarity between a correct answer in the input language and the response provided on each trial was calculated using normalised Damerau-Levenshtein edit distance ([Damerau 1964](#); [Levenshtein 1966](#)). Participants with an average distance of more than two edits per response, or greater than 20% of descriptions in which a word was omitted entirely or inserted were excluded.

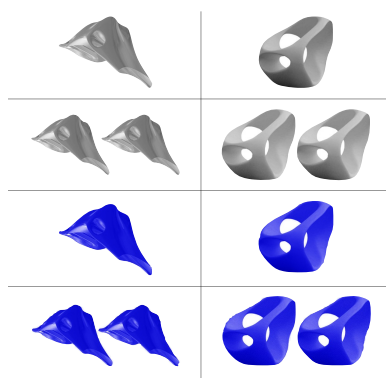


Figure 4.1: Visual stimuli used in the experiment. From top to bottom rows, pictures correspond to bare Nouns, noun phrases with a numeral modifier (Num Only), noun phrases with an adjective modifier (Adj Only), and noun phrases with both numeral and adjective modifiers (two-Modifier).

(Levenshtein 1966) between modifiers and nouns was held constant at $LD=4$; LD between modifier types (Num and Adj) was 5.

The two experimental conditions differed in the type of unconditioned probabilistic variation they used. In the Word Order condition, the two modifiers (Adj and Num) could appear either before or after the noun. Note that only isomorphic variants are possible (i.e., Adj and N are always adjacent), and the majority variants create a probabilistically harmonic system (i.e., both modifiers tend to appear post-nominally). In the Morphology condition, modifiers were consistently ordered after the noun, but each modifier had two distinct but related variants. These variants differed only in their second syllable². The respective probabilistic grammars for these two conditions are shown in Table 4.1.

Although the type of unconditioned variation in languages differs across experimental conditions (morphology vs. word order), imperatively system complexity is comparable. I assess a language's system complexity by its overall variability, computed using Shannon entropy; higher entropy corresponds to more variability, while lower entropy corresponds to more regu-

²This type of morphological variation could result from languages with rich inflectional morphology within NPs, such as Swedish where adjectives and determiners agree in gender and number with the head noun: e.g. *en grön stol* ('a green chair') and *min gröna stol* ('my green chair'), or *mitt gröna bord* ('my green table') and *mina gröna bord* ('my green tables').

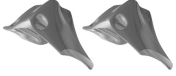


NP TYPE	MORPHOLOGY CONDITION	WORD ORDER CONDITION
NUM ONLY 	0.6 $NP \rightarrow \mathbf{N\ nefri}$	0.6 $NP \rightarrow \mathbf{N\ nefri}$
	0.4 $NP \rightarrow N\ nezno$	0.4 $NP \rightarrow nefri\ N$
ADJ ONLY 	0.6 $NP \rightarrow \mathbf{N\ kogla}$	0.6 $NP \rightarrow \mathbf{N\ kogla}$
	0.4 $NP \rightarrow N\ kospu$	0.4 $NP \rightarrow kogla\ N$
TWO-MODIFIER 	0.6 $NP \rightarrow \mathbf{N\ kogla\ nefri}$	0.6 $NP \rightarrow \mathbf{N\ kogla\ nefri}$
	0.13 $NP \rightarrow N\ kogla\ nezno$	0.13 $NP \rightarrow nefri\ kogla\ N$
	0.13 $NP \rightarrow N\ kospu\ nefri$	0.13 $NP \rightarrow nefri\ N\ kogla$
	0.13 $NP \rightarrow N\ kospu\ nezno$	0.13 $NP \rightarrow kogla\ N\ nefri$

Table 4.1: Probabilistic input languages in the Morphology and Word order conditions. Probabilities of occurrence for each variant within NP type are given to the left of the variant. Majority variants are highlighted in bold.

larity³. Entropy (H) of variant use in a given set of productions is given by

$$H(X) = - \sum_{i=1}^n P(x_i) \log_2 P(x_i), \quad (4.1)$$

where the sum is over the different variants, and $P(x_i)$ is the relative frequency of variant x_i in the set of productions, X . Following the probabilistic grammars described in Table 4.1, if the input language in the Word Order condition contains 30 instances of two-modifier phrases (i.e., 18 $N\ Adj\ Num$, 4 $Num\ Adj\ N$, 4 $Num\ N\ Adj$ and 4 $Adj\ N\ Num$), 10 instances of Num Only (i.e., 6 $N\ Num$ and 4 $Num\ N$) and 10 Adj Only (i.e., 6 $N\ Adj$ and 4 $Adj\ N$), the input language will have an overall entropy of 2.72 bits, Num Only and Adj Only would have an entropy of 0.97 each and two-modifier phrase would be defined by an entropy of 1.67 bits. If the language, on the other hand, was completely regular and thus only had one variant per phrase type, each phrase type would have a entropy of 0 bits and the overall language would be 1.37 bits. Provided the same number of instances of each phrase type, the input language in the Morphology condition

³Unlike other measures of complexity such as description length (see 1.2.3.1 in Chapter 1), simple Shannon entropy allows us to have a theory neutral measure of the system's complexity. Description length assigns more information content to terminal nodes than to combinatorial rules and thus presupposes that morphological variability (i.e. addition of terminal nodes) is more costly in information-theoretic terms than word order variability (i.e. addition of combinatorial rules).

results in the same exact entropy.

4.3.1.3 Procedure

Participants were taught a miniature artificial language through a video game interface developed in Python 2.7 (Peirce 2007, 2009; Pygame Community 2009). Each participant was trained and tested during a single 30-min-long experimental session. Participants sat in a sound-attenuated booth in front of a computer display and wore a headset through which the audio was played. All responses were entered using the keyboard. Before participants started the experiment, they were given detailed instructions by the experimenter on how to proceed during the experiment. Throughout the experiment, they were also given instructions by a “native speaker” of the artificial language, whose speech was synthetically generated using Apple’s text-to-speech software (OS version 10.10.3, speaker “Victoria”).

Phase 1, noun familiarisation Participants were first trained on the two nouns and their meanings during a block of 10 trials in which each noun appeared 5 times. The first six were simple exposure trials, and the remaining four were picture-selection comprehension trials. In exposure trials (here and throughout), participants were presented with a picture accompanied by its description in the language displayed both visually and aurally. In this phase, the picture was always a single object in grey-scale, and its description was a bare noun. The picture appeared first by itself (one second), then the text and audio were presented. After the audio finished, the text and the picture remained on display for two seconds. Participants were instructed to repeat the descriptions out loud. In picture-selection trials, participants were asked to select the picture (out of an array of four) which corresponded to a description in the language (again presented visually and aurally). The foils for the discrimination array in this phase consisted of the other object they had been trained on and two distractor objects (all in grey-scale and single). If the correct picture was selected, a correct-answer sound effect was played along with a display of the correct picture and description; if an incorrect picture was selected, a wrong-answer sound effect was played along with a display of the correct picture and description (in this case the audio of the description was also repeated).

Phase 2, one-modifier training In the second phase, participants were trained on noun phrases with a single modifier (either Adj or Num), here referred to as *one-modifier phrases*—which comprise *Adj Only* and *Num Only* phrases. Pictures corresponding to these phrases contained either a single object coloured blue or a grey-scale pair (see Figure 4.1). For each picture, a variant was selected probabilistically from the grammar corresponding to the participant's condition. Recall that both grammars contained a majority variant with an empirical probability $P=0.6$ and a minority variant with an empirical probability $P=0.4$ for each modifier type (see Table 4.1).

This phase included 40 trials in total, divided into two blocks of 20 trials; each block consisted of 15 exposure trials followed by five picture-selection trials. Participants saw each of the four one-modifier pictures five times per block (order randomised). The discrimination arrays on the picture-selection trials depended on the modifier category. If the target phrase included an Adj, the array contained each of the two objects in blue and greyscale, all singly; if the target phrase included a Num, the array contained each of the two objects singly and in pairs, all in grey-scale.

Phase 3, one-modifier testing In the third phase, participants were tested on their knowledge of one-modifier phrases in the language. They saw pictures as in the previous phase without accompanying text or audio and were asked to type in a description in the language. Participants described 20 pictures in total (five times each of the four different pictures), one at a time and in random order.

Phase 4, full training In the fourth phase, participants were trained on a mix of one-modifier and two-modifier phrases (i.e., Noun plus both Num and Adj). Two-modifier phrases were used to describe pairs of blue objects. For each picture, a variant was selected according to the grammar corresponding to the participant's condition. Recall that grammars contained a majority two-modifier phrase variant with an empirical $P=0.6$, and the three minority two-modifier phrase variants, each with an empirical probability $P=0.1\bar{3}$ (see 4.1).

This stage comprised 100 trials (20 Num Only, 20 Adj Only and 60 two-modifier phrases), divided into 4 blocks of 25 (15 exposure trials followed by 10 picture-selection trials). Participants saw each one-modifier picture 10 times and each two-modifier picture 30 times. In this

phase, the discrimination arrays in the picture-selection trials were of two types. In the first type, the array contained all four different pictures featuring the target object (i.e., grey-scale single object, blue single object, paired grey-scale objects, paired blue objects). In the second type, the array contained all four different pictures with the target modifier (if Adj, the bottom two rows in Figure 4.1; if Num, the top two rows in Figure 4.1).

Phase 5, full testing In the fifth phase participants were tested on their ability to produce all phrase types in the language. They saw all eight different pictures they had been trained on, and were asked to type in corresponding descriptions. They had to describe 52 pictures in total: 20 one-modifier pictures (10 Adj Only and 10 Num Only; 5 per object in each), 30 two-modifier pictures (15 per object) and the two bare object pictures (one each).

4.3.2 Analyses and results

4.3.2.1 Picture-selection

Across conditions, the average proportion of correct responses on picture-selection tasks during training was very high (Morphology condition; median = 0.95, MAD = 0.028; Word Order condition; median = 0.98, MAD = 0.018). Nevertheless, scores differed significantly between conditions (Mann-Whitney $U = 193$, $n_1 = n_2 = 26$, $p = 0.004$). This difference may have been driven by the larger lexicon in the Morphology condition (6 vs 4 lexical items).

4.3.2.2 Output variability

Entropy I assess participants' regularisation in production using Shannon entropy as described in section 4.3.1.2, equation 4.1. Entropy measures the variability of a given participant's productions; higher entropy correspond to more variability, while lower entropy corresponds to regularity. Using entropy instead of simply measuring the proportion of productions using the majority input variant allows us to capture regularisation behaviour more robustly (Ferdinand et al. 2017), in particular when more than two variants are available for a given meaning.

Figure 4.2 shows the entropy of participants' productions for both the Morphology and Word Order conditions. Analyses are run exclusively on participants' last set of production responses (phase 5, see section 4.3.1.3). Lexical items were corrected for typos before analysis.⁴

⁴ Typos were generally corrected to the closest vocabulary item. However I allowed for systematic innovations

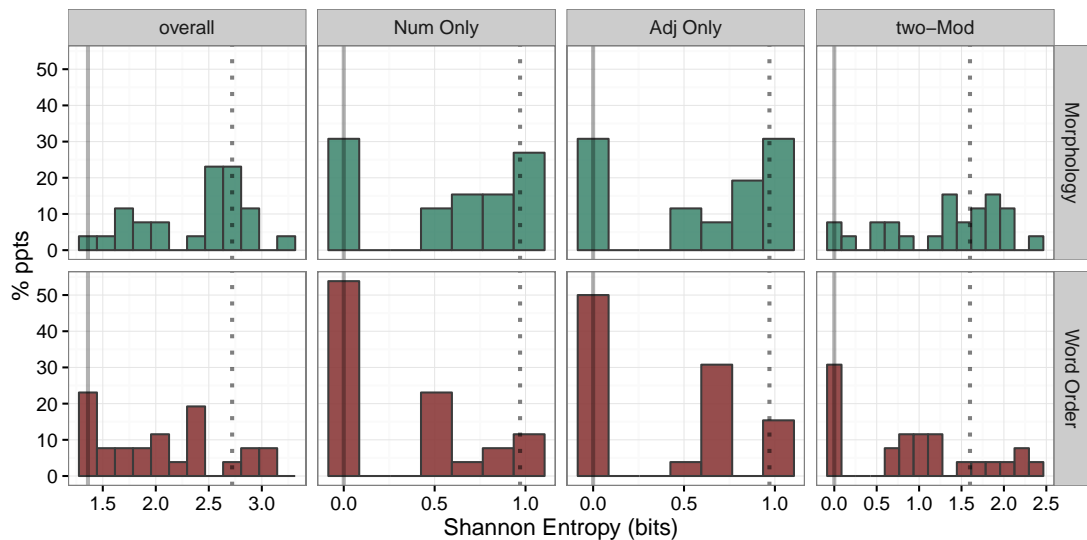


Figure 4.2: Entropy scores of participants' production systems. From top to bottom, scores for the Morphology (green) and Word Order (red) conditions. From left to right, entropies of participants' full production sets as well as entropies by NP type: one-modifier Num (Num Only), one-modifier Adj (Adj Only) and two modifier (two-Mod) NPs. Input entropy is indicated by a dashed vertical line. Minimum entropy scores are indicated by solid vertical lines. Minimum entropy is always 0 for each NP type in isolation but 1.37 for the overall system as it necessitates minimum 3 variants, one per NP type.

N1 and N2 were treated as the same Noun when the entropy of the phrases was calculated both in input and output systems, thus no variability was introduced by the correct use of the different nouns. An entropy of 0 corresponds to a set of productions that only contains one variant and therefore no variability whatsoever. Entropy lower and upper bounds vary according to the number of required and possible variants as well as to the number of production trials. The most regular language which is still expressive (i.e., contains a unique description for each picture) would consist of three different variants, one Num Only (e.g. *N nefri*), one Adj Only (e.g. *N kogla*) and one two-modifier (e.g. *N kogla nefri*). The final production phase consisted of 50 trials (excluding the 2 Noun trials), divided up into 20 one-modifier trials (half Num and half Adj) and 30 two-modifier trials: the entropy lower bound for the *overall* language is thus 1.37 bits (represented as a solid vertical line in Figure 4.2). The input *overall* entropy for the same number of trials would be 2.72 bits (represented as a dotted vertical line in Figure 4.2). Note that whereas the input entropy and the output entropy lower bound is held constant, output

by participants. For example, in the Morphology condition, only one-off typos or consistent misspelling of lexical items were corrected, so that additional variants introduced systematically by a given participant were retained. Similarly, in the Word Order condition, production of two-modifier phrase word orders which were not present in the input were not corrected. These additional variations introduced by participants could therefore increase their entropy scores.

entropy upper bounds will increase if participants introduce additional variants.

Figure 4.2 shows entropy for the different phrase types separately: one-modifier Num (Num Only), one-modifier Adj (Adj Only), and two-modifier (two-Mod) phrases. As the most regular system would contain a single variant per NP type, the minimum entropy for the set of production for a given NP type individually is 0. A visual inspection of the output entropy across the Morphology and Word Order conditions suggests that many participants did not reproduce the full variability of the input languages; their productions are generally more regular than the input (i.e., entropy is lower). Moreover, the modes in output entropy in Figure 4.2 hint at higher regularity in the Word Order condition.

I used the *stats* and *lme4* packages developed in R (Bates et al. 2015; R Core Team 2015) to run a linear mixed-effects model predicting regularisation (entropy) by Condition (two levels: Morphology as reference, and Word Order), NP Type (reverse Helmert coded with 3 ordered levels: Num Only, Adj Only and two-Mod) and System (two levels: Input as reference, and Output). The model included main effects and all interactions. The contrast coding used means that the model compares Adj Only to Num Only, and two-Mod to the mean of those. In addition to these fixed effects, random intercepts for Subject as well as by-Subject random slopes for the effects of NP Type and System were also included. P-values were obtained through the *lmerTest* package (Kuznetsova et al. 2015), using Satterthwaite approximation (Satterthwaite 1946). Results show a significant effect of System ($\beta = -0.346, SE = 0.085, p < 0.001$), suggesting that participants do indeed regularise the input in their output productions. I also found a significant interaction between System and Condition ($\beta = -0.284, SE = 0.119, p = 0.021$), suggesting that participants regularise their input significantly more in the Word Order condition. I found the expected effect of higher input entropy in two-Mod phrases ($\beta = 0.21, SE = 0.024, p = < 0.001$), but no significant interactions between NP Type and System (largest: $\beta = 0.027, SE = 0.028, p = 0.324$) or between NP Type, System and Condition (largest: $\beta = -0.041, SE = 0.039, p = 0.299$). These results suggest that participants regularised input languages across conditions and NP types, and that participants in the Word Order condition regularised more than those in the Morphology condition on average.

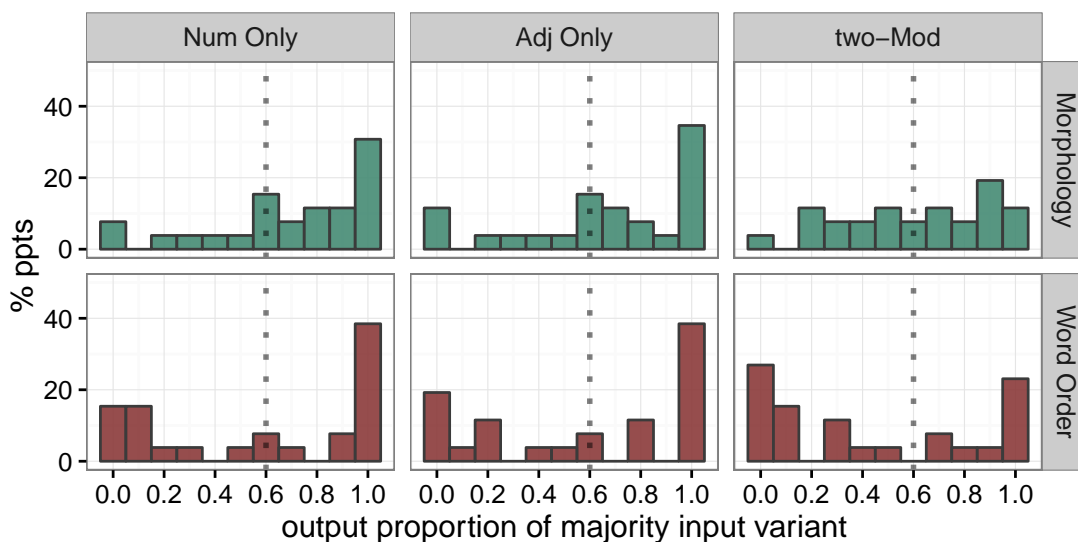


Figure 4.3: Output proportion of the majority input variant for each of the NP types in the Morphology (green) and Word Order (red) conditions.

Variant production The previous analyses have shown that participants regularised input variability in both conditions, but more so in the Word Order condition. However, this reduction in variability could be due to over-production of majority input variants or it could reflect over-production of minority variants. Figure 4.3 shows the proportion of majority input variants produced by participants across conditions. I ran a logistic mixed-effects model to evaluate the proportion of the majority input variant in the output compared to the input languages across conditions. Responses in production were dummy coded according to the presence or absence of the majority input variants in each trial. The fixed effects structure was the same as in the previous model. As random effects, I included random intercepts for Subject as well as by-Subject random slopes for the effect of NP Type. I found a significant effect of Output ($\beta = 0.289, SE = 0.096, p = 0.0027$) as well as a significant interaction between Output and Word Order ($\beta = -0.599, SE = 0.137, p < 0.001$), suggesting that, while participants in the Morphology condition over-produce the input majority variants, those in the Word Order condition do not. I also found a significant effect of the interaction between two-Mod and Output ($\beta = -0.123, SE = 0.057, p = 0.032$), and no significant interaction between two-Mod, Output and Word Order ($\beta = -0.1, SE = 0.082, p = 0.223$), suggesting that participant across conditions produced majority input variants less frequently for two-modifier phrases than for one-modifier phrases.

A visual inspection of Figure 4.3 suggests that all distributions in the Word Order condition

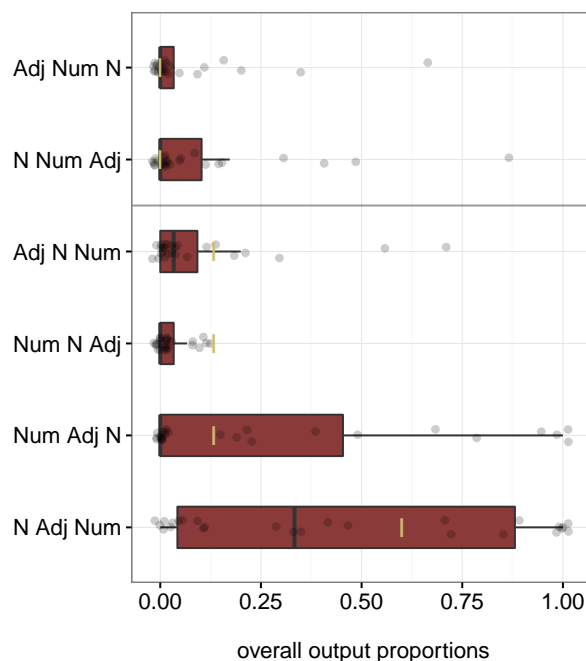


Figure 4.4: Overall output proportions for each two-modifier variant produced in the Word Order condition. Seen (bottom) and unseen (top) variants during training are divided by a solid grey line. Vertical yellow lines indicate input proportions.

are bimodal, with modes of the distributions of majority variant use at $P \leq 0.1$ and $P > 0.9$ across phrase types, suggesting two opposite trends amongst participants: one towards the over-production of the majority input word order variants and another towards under-production of majority variants. Participants under-producing the majority word order variant in *one-modifier phrases* are necessarily producing modifiers pre-nominally. Figure 4.4 shows the overall proportions of the variants produced for *two-modifier phrases* by all participants. The input proportions are represented by the yellow vertical lines. The word order produced the most is the majority input variant N Adj Num, but it is regularised only by a minority of participants. Although the three remaining word order variants were equally frequent in the input language, the opposite harmonic order Num Adj N order is produced more frequently by participants (though only by a minority as indicated by the median value 0). In total, only 30% of participants produced both harmonic variants (and only 19% produced both variants more than once), suggesting that although harmonic orders are preferred overall, they do not generally coexist within the productions of a single participant. I ran a mixed-effects logistic regression to explore the difference between proportions of the Num Adj N variant in input and output languages. The model included System (Input as reference, and Output) as the only

fixed effect. Random intercepts for Subject as well as a by-Subject random slope for the effect of System were also included. Results show that the Num Adj N is produced significantly less in output languages than in the input ($\beta = -7.641, SE = 1.984, p < 0.001$). Only a minority of participants overproduced this variant, the majority of participants in fact under-produced it. On top of the observed preference for harmonic order, these results suggest (or at least do not contradict) a tendency to avoid the coexistence of two opposite N-peripheral variants, i.e., N Adj Num and Num Adj N.

4.3.3 Interim discussion

Previous work has suggested a link between the apparent rarity of unconditioned variation in language, and biases active during learning. Adult learners tend to regularise unconditioned variation, particularly when the systems they are learning are relatively complex or contain dispreferred variants (Culbertson et al. 2012; Hudson Kam & Newport 2009; K. Smith & Wonnacott 2010). Consistent with these previous findings, participants in Experiment 5 regularised systems of probabilistic unconditioned variation in both morphology as well as in word order. Although the majority of participants remained probabilistic users (only a minority used a completely deterministic system), most participants in both conditions regularised their input by eliminating variants and/or by increasing the frequency of a variant and reducing others.

Yet, despite their similarity in terms of overall system complexity, regularisation was not equal across the Word Order and Morphology conditions. Participants regularised more when the input involved variation in word order compared to variation in morphology (represented by variation in word endings). This difference is in the opposite direction with regard to what I predicted on the basis of previous literature on language learning and formation (e.g., pidginisation). In this literature, the suggestion is that complex morphological variation is more likely to be regularised or simplified, while syntactic variation is more likely to be retained (Anderssen et al. 2010; Bichakjian 1988; Drechsel 1981). However, a close analysis of variant usage in the Word Order condition suggests that the difference in strength of regularisation behaviour between conditions might reflect additional biases that are relevant only for word order. First, a bias in favour of harmonic variants (N Adj Num and Num Adj N) in general, and against opposite N-peripheral patterns in the same grammar. These patterns are maximally dissimilar to one another in terms of the surface order of the elements involved, therefore it is perhaps

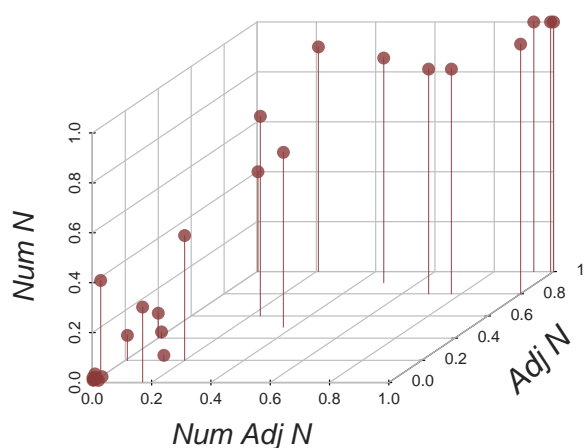


Figure 4.5: Relationship between output proportion of pre-nominal word orders across phrase types. We observe that most participants do not use pre-nominal word orders at all but that the more they use pre-nominal two-modifier phrases, the more they use pre-nominal one-modifier phrases.

not surprising that participants would avoid using both. Second, a bias favouring Num Adj N word order coming from participants' native language, English (Weber et al. 2016). Over- or under-use of this order in two-modifier phrases caused by these two factors may have led some participants to accordingly over- and under-using pre-nominal modification in one-modifier phrases as well. A visual inspection of Figure 4.5 suggests a positive relationship between the proportion of pre-nominal orders in one-modifier and the proportion of two-modifier phrases⁵. These biases together will contribute to additional reduction in the amount of output variation in this condition relative to the Morphology condition. To avoid co-occurrence of the two favoured harmonic orders, a participant may overproduce one and under-produce the other. If either order is overproduced in two-modifier phrases, this may lead to reduced variation in the order of one-modifier phrases as well. To minimise these effects, I ran a second experiment, in which unconditioned probabilistic variation in word order was present, but the input language did not contain the English word order (and thus did not include both N-peripheral patterns).

⁵Additionally, a logistic regression mixed-effects model shows that the proportions of pre-nominal two-modifier phrases do not differ from those predicted from the product of Num N and Adj N productions within one participant's use ($\beta = -0.198, SE = 0.193, p = 0.303$), which suggests a non-trivial relationship between pre-nominal modification in one-modifier and two-modifier variants.



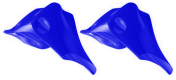
NP TYPE	WORD ORDER CONDITION	NO L1 WORD ORDER CONDITION
NUM ONLY 	0.6 <i>NP</i> → N nefri	0.6 <i>NP</i> → N nefri
	0.4 <i>NP</i> → nefri N	0.4 <i>NP</i> → nefri N
ADJ ONLY 	0.6 <i>NP</i> → N kogla	0.6 <i>NP</i> → N kogla
	0.4 <i>NP</i> → kogla N	0.4 <i>NP</i> → kogla N
TWO-MOD 	0.6 <i>NP</i> → N kogla nefri	0.6 <i>NP</i> → N kogla nefri
	0.1 $\bar{3}$ <i>NP</i> → nefri kogla N	0.1 $\bar{3}$ <i>NP</i> → N nefri kogla
	0.1 $\bar{3}$ <i>NP</i> → nefri N kogla	0.1 $\bar{3}$ <i>NP</i> → nefri N kogla
	0.1 $\bar{3}$ <i>NP</i> → kogla N nefri	0.1 $\bar{3}$ <i>NP</i> → kogla N nefri

Table 4.2: Probabilistic input language in the NoL1 Word order condition contrasted with the Word Order condition in Experiment 5. Changes in the variant set are indicated with boxes. Probabilities of occurrence for each variant within NP type are given to the left of each variant. Majority variants are highlighted in bold.

4.4 Experiment 6: the effect of the L1 two-modifier variant on regularisation behaviour

Experiment 6 follows the same design as the Word Order condition described in Experiment 5, with a single change: the set of two-modifier input variants. As illustrated in Table 4.2, I swapped the two-modifier Num Adj N variant for N Num Adj, maintaining the number of harmonic orders but eliminating the L1 variant and the presence of opposite N-peripheral patterns within two-modifier phrases. These changes should mitigate the effect of L1 transfer and increase the coexistence of harmonic patterns (i.e., N Adj Num and N Num Adj) in production. If the effects outlined above were indeed contributing to the difference between conditions in Experiment 5, then I expect these changes to lead to decreased regularisation of word order. For descriptiveness, I call Experiment 6 the *NoL1 Word Order* condition.

4.4.1 Participants

Twenty-six native-English speakers (10 male, 16 female; aged between 18 and 35, mean=24.8) were recruited via the University of Edinburgh's Careers Service advertisement database. Par-

participants received £6. The data from a further two participants were excluded on the basis of failure to adequately learn the language. This was determined as per Experiment 5 (see section 4.3.1.1).

4.4.2 Analyses and results

4.4.2.1 Comprehension

As in Experiment 5, proportion of correct selections on picture-selection tasks during training were high in the NoL1 Word Order condition (median = 0.97, MAD = 0.027). Accuracy scores were slightly higher than in the Morphology condition of Experiment 5 (Mann-Whitney $U = 223.5$, $n_1 = n_2 = 26$, $p = 0.017$, corrected- $p = 0.05$), but similar to those in the Word Order condition (Mann-Whitney $U = 326.5$, $n_1 = n_2 = 26$, $p = 0.41$).

4.4.2.2 Output variability

Entropy Entropy scores obtained in the NoL1 Word Order condition are shown in Figure 4.6, contrasted with those in the Morphology and Word Order conditions in Experiment 5. I ran a linear mixed effects model to explore the effect of condition on regularisation behaviour now including NoL1 Word Order as well as Morphology and Word Order in Experiment 5. As fixed effects I entered NP Type (reverse Helmert coded with 3 ordered levels: Num Only, Adj Only and two-Mod NPs), Condition (reverse Helmert coded with 3 ordered levels: Morphology, NoL1 Word Order and Word Order) and System (two levels: Input as reference, and Output), as well as all interactions. Condition was reverse Helmert coded so that NoL1 Word Order was directly compared to the Morphology condition from Experiment 5, and the Word Order condition was compared to the average of those. I also entered random intercepts for Subject as well as by-Subject random slopes for the effect of NP Type and System. I found a significant effect of System ($\beta = -0.483$, $SE = 0.051$, $p < 0.001$) and a significant interaction between Word Order and System ($\beta = -0.073$, $SE = 0.036$, $p = 0.046$), ratifying regularisation behaviour across conditions and its higher strength in the Word Order condition. However, I did not find a significant interaction between NoL1 Word Order and System ($\beta = -0.063$, $SE = 0.063$, $p = 0.317$), suggesting that participants in the Morphology and the NoL1 Word Order conditions regularised their input to similar degrees. I did not find signifi-

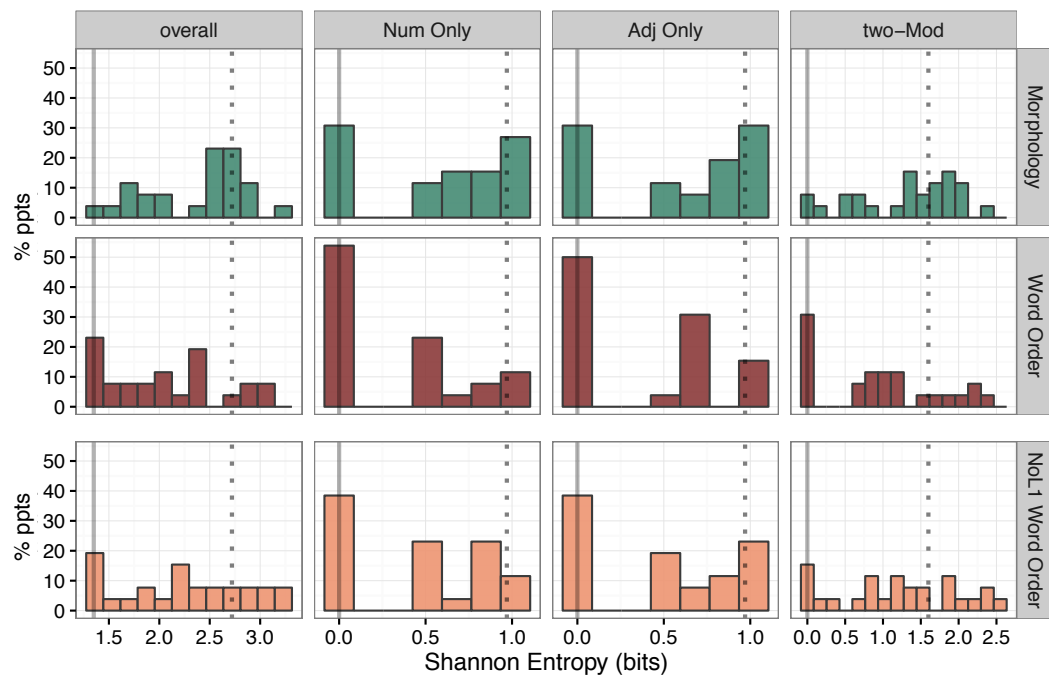


Figure 4.6: Entropy scores of participants' production systems. From top to bottom, scores for the Morphology (green) and Word Order (red) conditions in Experiment 5 and for the NoL1 Word Order condition (orange) in Experiment 6. From left to right, entropy scores of participants' full production sets as well as entropies by NP type: one-modifier Num (Num Only), one-modifier Adj (Adj Only), and two modifier (two-Mod) NPs. Input Entropy scores are indicated by dashed vertical lines. Minimum entropy scores are indicated by solid vertical lines. Minimum entropy is always 0 for each NP type in isolation but 1.37 for the overall system as it necessitates a minimum of three variants, one per NP type.

cant interactions between NP Type and System (largest: $\beta = 0.016, SE = 0.015, p = 0.288$) or between NP Type, System and Condition (largest: $\beta = -0.015, SE = 0.011, p = 0.168$). These results suggest that participants regularised their input languages significantly across conditions and NP types, and that whilst participants in the Word Order condition are regularising more than those in the Morphology condition, the latter and participants in the NoL1 Word Order condition regularise their input to similar degrees. Excluding the Num Adj N variant in the input language thus eliminated the difference between levels.

VARIANT PRODUCTION Figure 4.7 visualises the proportion of majority input variants produced for each phrase type in the NoL1 Word Order in comparison with both conditions in Experiment 5: we observe a tendency towards overproduction of the majority variant across one-modifier NPs but not in two-modifier NPs. As in Experiment 5, I ran a logistic mixed-effects model comparing the proportion of majority input variants in input and output languages, but this across all conditions in Experiments 5 and 6: I used simple contrast to compare Morphology and Word Order to NoL1 Word Order. Thus as fixed effects I introduced Condition (NoL1 Word Order as reference), NP Type (reversed Helmert coded as per previous models) and System (output vs. input), as well as all their interactions. As random effects, I included intercepts for Subject as well as by-Subject slope for the effect of NP Type. Results suggest that participants in the NoL1 Word Order condition overproduce the majority variant in one-modifier phrases ($\beta = 0.212, SE = 0.098, p = 0.031$) but not in two-modifier phrases ($\beta = -0.324, SE = 0.058, p < 0.001$). I found a significant interaction between the production of two-Mod majority variants and Morphology ($\beta = 0.201, SE = 0.082, p = 0.014$) but not with any other NP types (largest: $\beta = 0.112, p = 0.559$), ratifying that participants in Morphology overproduce majority variants across NP types. Moreover, I found a significant interaction between production of majority variants and Word Order ($\beta = -0.522, SE = 0.138, p < 0.001$) and no three-way interaction (largest: $\beta = 0.1, p = 0.222$), also a ratification of the lack of overproduction of majority input variants in Word Order.

Figure 4.8 shows the overall output proportions of not only the majority input variant but all the different variants produced for two-modifier NPs in the NoL1 Word Order condition. We observe that, although not over-produced on average, N Adj Num (the majority input variant) is still the most frequent on average. Within minority input variants, the harmonic N Num Adj

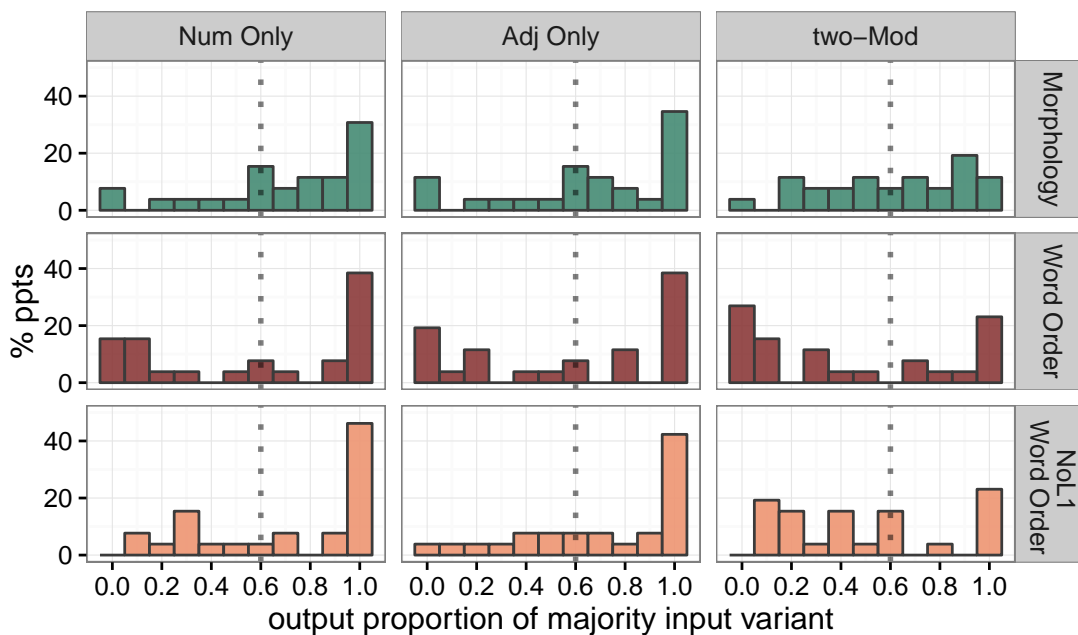


Figure 4.7: Output proportion of the majority input variant for each of the NP types in NoL1 Word Order (orange), compared to Morphology (green) and Word Order (red).

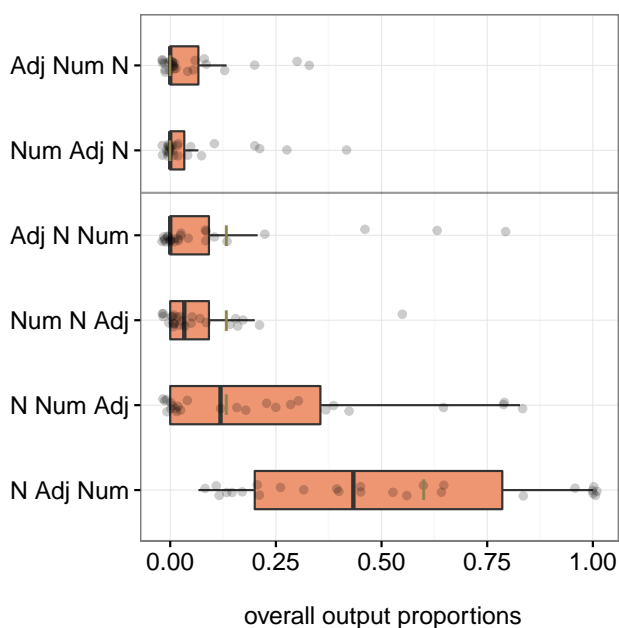


Figure 4.8: (a) Overall output proportions for each two-modifier variant produced in the NoL1 Word Order condition. Seen (bottom) and unseen (top) variants during training are divided by a solid grey line. Vertical brown lines indicate input proportions.

word order is observably most frequent. Unlike in the previous Word Order condition where the two input harmonic patterns were very different from one another (i.e., Num Adj N and N Adj Num), in the NoL1 Word Order condition, 65% of participants produced systems with both N Adj Num and N Num Adj harmonic variants. I ran a logistic mixed-effects regression model to test the difference between the proportions of N Num Adj variants in input and output languages. I entered System (two levels: Input as reference, and Output) as the only fixed effect. Random intercepts for Subject and by-Subject random slopes for the effect of System were included. Results show that the proportion of N Num Adj variants in the output languages was not significantly different from the input proportion ($\beta=-0.594$, $SE=0.546$, $z=-1.086$, $p=0.277$). These results suggest that unlike in Morphology or Word Order, the regularisation of two-modifier systems in the NoL1 Word Order condition is not driven by general tendencies for over- or under-production of specific variants.

4.5 Discussion: the effect of linguistic level on regularisation in isolate production

The experimental results reveal regularisation in the production of unconditioned variation in morphology and word order, in line with an overarching simplicity bias shown to be at play in language learning and use (Culbertson & Kirby 2016). Moreover, regularisation behaviour is of similar strength between linguistic levels given input languages with comparable initial complexity. In Experiment 5 I found higher levels of regularisation in word order than in morphology; however this was determined to be due to the specific set of variants present in the input language. When both harmonic pre-nominal and post-nominal two-modifier variants were included, some participants over-used the English-like Num Adj N variant. Further, the coexistence of these two quite distinct variants in a single participant's productions was rare, leading to decreased variability in this condition overall. In Experiment 6, I showed that eliminating opposite N-peripheral orders from the set of two-modifier variants (by swapping Num Adj N for N Num Adj) eliminated the difference in regularisation between levels. These findings therefore do not support global level-specific biases as an explanation for the asymmetry between the simplification of morphological and word order hinted at in the literature on language learning and pidgin formation (Bichakjian 1988; Drechsel 1981; Good 2015). Instead,

this study suggests that comparably complex linguistic systems simplify to similar degrees and thus any asymmetry is more likely due to differing complexity of linguistic paradigms or other features of the contact languages (Ansaldo et al. 2007; Mufwene 2008).

To summarise, results so far suggest similar strengths of regularisation behaviour in production across linguistic levels. In particular, when input languages have comparable initial complexities and there are no specific biases targeting alternative structures available at only one level (in this case word order), no difference in regularisation was found. However, Experiments 5 and 6 do not necessarily reflect the context within which languages—including newly formed pidgins—are learned and used. Language production typically takes place in a context of communicative interaction and so far I have only tested production of variants without an explicitly communicative task. Some studies on language emergence (i.e., pidgin/creole formation) have contemplated grammatical simplification as the product of the reduction of features in source languages which are “incidental to basic communication” (McWhorter 2001) (for similar views in L2 acquisition more broadly, see Klein & Perdue 1997). These studies assume thus that language users would only produce the minimum grammatical information required to convey meanings, which will be dependent on the communicative context. Under this view, language users would thus omit any uninformative variability in their productions as a direct consequence of a communicative need. Indeed, previous experimental studies have indicated that an additional source of regularisation may come from alignment in communication, raising the possibility that differences across linguistic levels may appear in communicative contexts. For example, Fehér, Wonnacott, and Smith (2016) report results from an artificial language learning task in which participants exposed to unconditioned variation must interact with other learners after training. In their study, pairs of participants were trained on a linguistic system with variable word order (two different word orders, each appearing 50% of time) and later were asked to recall it in isolation, use it to communicate with an interlocutor, and finally recall it in isolation once again. They find increased regularisation during the communication phase, suggesting that a mechanism such as reciprocal priming or alignment might contribute to explaining the tendency for avoiding unconditioned variation in natural language. In Experiment 7 I assess the effect of linguistic level on regularisation behaviour in a more naturalistic context for production including communicative interaction.

4.6 Experiment 7: regularisation behaviour across linguistic levels in communicative interaction

In Experiment 7 I utilise the methodology developed by Fehér, Wonnacott, and Smith (2016) to incorporate communicative interaction into the experimental paradigm. As in the previous experiments, adult learners are exposed to a miniature artificial language featuring an inconsistent mixture of synonymous variants. However, this time all production takes place in the context of dyadic communication.

4.6.1 Materials and methods

4.6.1.1 Participants

Forty-eight English-native speakers (13 male, 35 female; aged between 18 and 26, mean 21.8) were recruited from the University of Edinburgh's Careers Service database of vacancies. Each was compensated £6. Twenty-four participants (12 pairs) were assigned to the Morphology Dyads condition, and the other 24 (12 pairs) to the Word Order Dyads condition. No further participants were excluded in this experiment.

4.6.1.2 Input languages

The input languages used for the Morphology Dyads and the Word Order Dyads conditions were the same ones used for the Morphology condition in Experiment 5 (see Table 4.1) and the NoL1 Word Order condition in Experiment 6 (see Table 4.2) respectively.

4.6.1.3 Communicative interaction

The experimental procedure was the same as described in section 4.3.1.3 for the previous experiments with the only modification being the introduction of dyadic communicative interaction during all testing stages, i.e., experimental phases 3 and 5 (see section 4.3.1.3). The instructions provided were changed accordingly: participants were told that they were going to be taught an artificial language which they had to learn to later communicate with their partners in testing phases. In Phase 3 (one-modifier testing), participants in a dyad took turns describing one-modifier pictures to their partner, who then had to interpret this description by choosing

the corresponding picture from a set. They completed 20 trials total (10 trials per one-modifier type (Num Only and Adj Only), 5 per object). In Phase 5 (full testing), participants in a dyad took turns describing and interpreting descriptions for 52 trials. These 52 trials comprised the following set of pictures (shown in random order): 10 Adj Only (5 per object), 10 Num Only (5 per object), 30 two-Mod (15 per object) and two noun pictures (one per object, singly and in grey-scale).

At each testing trial, one participant was assigned the role of “sender” and the other that of the “receiver”. The sender typed in a description for the picture appearing in their screen. The description was sent to and received by the other participant in the dyad. The receiver saw the description in the middle of the screen surrounded by an array of four different pictures (chosen as described in section 4.3.1.3 for picture selection trials). The receiver had to select the picture they thought their partner wanted to convey. After the receiver selected a picture, both participants were given feedback on their communicative success: a green screen was displayed after the receiver correctly selected the picture the sender described, and a red screen when the receiver selected a different picture. Green screens were accompanied by a correct-answer sound effect and red screens by a wrong-answer sound effect. The roles of sender and receiver were swapped after each trial.

4.6.2 Analyses and results

4.6.2.1 Picture-selection

As in the previous two experiment, participants achieved high accuracy in picture-selection tasks across conditions (Morphology Dyads condition; median = 0.967, MAD = 0.011; Word Order Dyads condition; median = 0.972, MAD = 0.017). In this case, accuracy scores did not differ significantly between conditions (Mann-Whitney $U=241.5$, $n_1=n_2=24$, $p=0.169$). During testing, communicative accuracy scores (i.e., the proportion of successful communication trials) were also equally high in both conditions (Morphology Dyads: median = 0.925, MAD = 0.075; Word Order Dyads: median = 0.95, MAD = 0.052; Mann-Whitney $U = 62.5$, $n_1 = n_2 = 12$, $p = 0.297$).

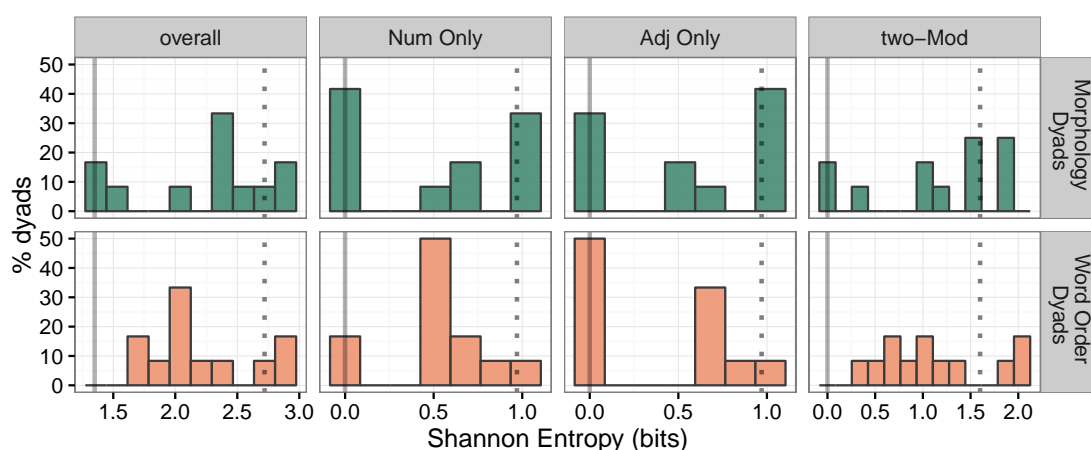


Figure 4.9: Entropy scores of the systems produced by dyads. Scores are shown for the Morphology Dyads (green) and Word Order Dyads (orange) conditions. From left to right, entropies of full production sets as well as entropies by NP type: one-modifier Num (Num Only), one-modifier Adj (Adj Only), and two-modifier (two-Mod) NPs. Input Entropy scores are indicated by dashed vertical lines. Minimum Entropy scores possible are indicated by solid vertical lines.

4.6.2.2 Output variability

Entropy Figure 4.9 shows the Shannon entropy of participants' production systems in Morphology Dyads and Word Order Dyads. I show the entropy of each dyad's set of productions ($N = 12$ for each condition). A visual inspection of the entropy scores obtained suggests that most dyads failed to reproduce the full variability of the input languages. I ran a linear mixed-effects model to test the effects of linguistic level (i.e., word order vs. morphology) and communicative interaction on regularisation behaviour (DV: Shannon entropy). I compare the Morphology in Experiment 5 and NoL1 Word Order in Experiment 6 to Morphology Dyads and Word Order Dyads respectively. I entered four fixed effects and their interactions into the model: Linguistic Level (word order vs. morphology), Experiment (Communicative interaction as reference, and Isolate production), NP Type (Num Only, Adj Only and two-Mod), and System (Input as reference, and Output). As random effects, I entered an intercept for Subject (here a dyad instead of an isolate) and by-Subject random slopes for the effect of System and NP Type. I used reverse Helmert contrasts on NP Type and Linguistic Level. For NP Type, the model compares Adj Only to Num Only and then two-Mod to the average of those. For linguistic level, Word Order is compared to Morphology⁶. Results show that dyads also regularise their input significantly on average across linguistic levels and NP Type

⁶This contrast was simply used to overcome problematic collinearity.

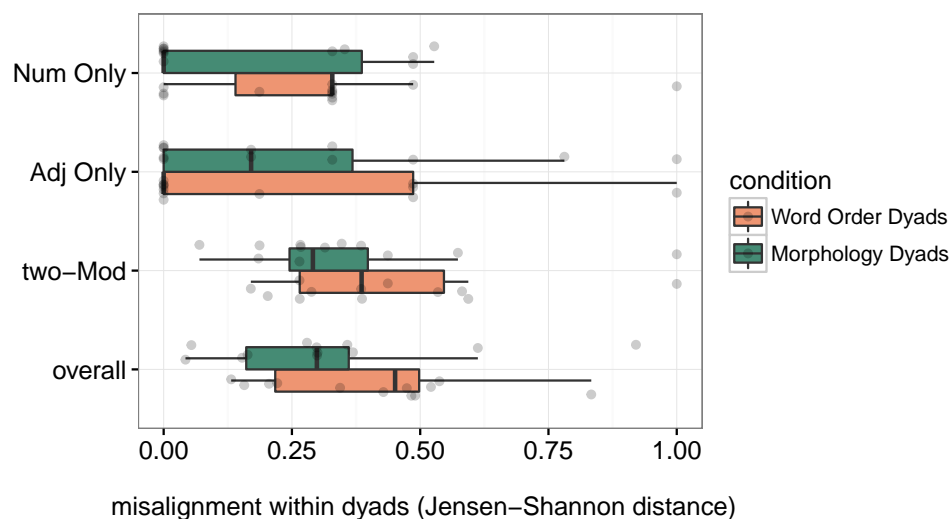


Figure 4.10: Misalignment between participants in a dyad. From bottom to top, I provide the Jensen-Shannon distances for the overall systems, the two-Mod NPs alone, Adj Only NPs and Num Only NPs. Individual data points from each dyad as well as a boxplot summarising the data are provided (green and orange for Morphology Dyads and Word Order Dyads respectively). We observe that dyads are not as closely aligned as expected if participants were matching each other's behaviour, specially in two-Mod productions, where the distances between variant production are highest.

($\beta = -0.480, SE = 0.090, p < 0.001$). The level of regularisation did not differ significantly between linguistic levels ($\beta = -0.025, SE = 0.09, p = 0.782$) nor between production in isolation and during communicative interaction ($\beta = 0.071, SE = 0.109, p = 0.517$).

Alignment The results above contrast with those reported in [Fehér, Wonnacott, and Smith \(2016\)](#) in that mine suggest similar degrees of regularisation behaviour across isolate production and communicative interaction. This difference could be caused by the combination of two factors. Firstly, input languages in this study are relatively more complex systems in terms of the variants available: eight different variants across three phrase types. Indeed, participants in Experiments 5 and 6 already regularise these systems when they produce phrases in isolation, and therefore the level of regularisation may not increase further once communication is added. The second factor is the lack of perfect alignment during interaction, i.e., the misalignment between two interlocutors' productions within a dyad. If participants within a dyad regularise the input in different ways and therefore the distance between their production systems is substantial, the admixture of the two systems will result in higher variability. I use the Jensen-Shannon distance ([Endres & Schindelin 2003](#)) to measure the distance between the

frequency distributions over variants in the production systems of each of the two participants in a dyad. The Jensen-Shannon distance is the square root of the Jensen-Shannon divergence, which is a symmetrised version of the more general Kullback-Leibler divergence metric. Let P and Q be two probability distributions. Kullback-Leibler divergence is defined as:

$$D_{KL}(P||Q) = \sum_i P(i) \log_2 \frac{P(i)}{Q_i}. \quad (4.2)$$

I then symmetrise this expression and take the square root to obtain the Jensen-Shannon distance, given by

$$JSD(P||Q) = \sqrt{\frac{D_{KL}(P||M) + D_{KL}(Q||M)}{2}}, \quad (4.3)$$

where $M = (P + Q)/2$. Figure 4.10 shows the Jensen-Shannon distance of all dyads—for whole systems as well as for each NP Type. We observe that dyads are not perfectly aligned in most cases, specially for two-Mod productions, where the use of different variants is often quite distinct. I ran a linear mixed-effects model to explore misalignment (DV: Jensen-Shannon distance) across conditions and across the different types of Phrase Types. As fixed effects I entered Condition (Morphology Dyads and Word Order Dyads) and NP Type (Num Only, Adj Only and two-Mod) with an interaction term. I used reverse Helmert contrasts across fixed effects so the model's intercept is the average across conditions and across NP types. For Condition, Word Order is compared to Morphology; and for NP Type, Adj Only is compared to Num Only, and two-Mod is compared to the average of those. As random effects, I entered intercepts for Subject. The model intercept shows that the Jensen-Shannon distance is significantly different from $JSD(P||Q) = 0$ (i.e., perfect alignment) on average across conditions and Phrase Types ($\beta = 0.294, SE = 0.043, p < 0.001$). Moreover, I found a significant effect of two-Mod ($\beta = 0.049, SE = 0.016, p = 0.004$), suggesting that misalignment between participants' productions is higher for two-modifier than for one-modifier NPs (as observed in Figure 4.10). Results did not show any differences between conditions (largest: $\beta = -0.043, SE = 0.028, p = 0.128$).

4.6.3 Discussion: the effect of linguistic level on regularisation behaviour during communicative interaction

Previous work has framed communicative interaction as a potential mechanism for regularisation of probabilistic unconditioned variation (Fehér et al. 2017; Fehér, Wonnacott, & Smith 2016). I therefore incorporated a communicative task into the experiment in order to determine whether differences in regularisation across linguistic levels might emerge in this context. In particular, participants had to use phrases in the language to describe pictures to a partner and had to interpret phrases produced by that partner. Thus, an increase in regularisation was possible through priming or alignment in this experiment. However, results revealed no significant additional regularisation compared to production in isolation in Experiments 5 and 6. More importantly, no difference across linguistic levels was revealed.

This experiment concludes that the communicative context does not straightforwardly lead to stronger regularisation behaviour. The discrepancy between my results and those in Fehér, Wonnacott, and Smith (2016) might stem from the differences in the complexity of the systems of variation. Unlike those in Fehér, Wonnacott, and Smith (2016) and those in a later study (Fehér et al. 2017), results from Experiment 5 and 6 suggest that participants regularise during isolate production. Therefore, in combination with the relatively significant misalignment between pairs of participants in Experiment 7, I find similar degrees of regularisation across experiments.

First and foremost, I conclude the communicative context does not seem to provide an alternative explanation for differences in regularisation across linguistic levels either. These results thus suggest that simplification in contexts of pidgin formation or L2 production more broadly cannot be reduced to the communicative context (cf. Klein & Perdue 1997; McWhorter 2001) and neither can the differences between linguistic levels.

4.7 Encoding

The majority of experiments on regularisation illustrate its effects using production (similar to the production tests used in the experiments in this study); production in these studies is taken to be a reflection of what learners encode during learning. However, recent studies have suggested the possibility that regularisation indeed occurs mostly during production, with variation

actually being learned or encoded with relatively high accuracy. For example, [Ferdinand et al. \(2017\)](#) show that adult learners regularise across a number of conditions, from a very simple task involving a single object with two alternating labels, to a much more difficult task with many objects and alternating labels and thus many more frequencies to track. In fact [Ferdinand et al. \(2017\)](#) show that participants are able to accurately report input proportions, even when they exhibit regularisation in their productions. In line with this finding, [Perfors \(2012a, 2012b, 2016\)](#) shows that, although increasing working memory load during training does not lead to regularisation ([Perfors & Burns 2010](#)), perception of variation does. The more random participants perceived the system to be, the more they regularised input variation. Furthermore, [Fehér, Ritt, et al. \(2016\)](#) reveal that the same participants regularise their input significantly during language use in communication but they are nevertheless able to reproduce the input variability during isolate production. Taken together, these results suggest the possibility that a regularisation bias may be found specifically in production, rather than in learning or encoding. Thus in order to shed light onto the cognitive roots of regularisation behaviour in the experiments reported above, I investigate not just whether participant's *produce* regularised systems, but whether regularisation is found in encoding as well. I do this through grammaticality judgements and post-experimental questionnaires in which participants are asked to estimate the relative frequencies of variants in the input languages.

4.7.1 Materials and methods

All participants in Experiments 5, 6 and 7 were asked to complete a grammaticality judgement task and to fill in an additional questionnaire (described below), which was displayed on the monitor via a web browser.

4.7.1.1 Grammaticality judgement

After participants completed all training and testing phases, they were asked to provide grammaticality judgements for phrases in the language. Participants saw a picture and a description and were asked to accept or reject the description as part of the language. They saw 24 picture-phrase pairs, 12 one-modifier trials (six Num Only and six Adj Only) and 12 two-modifier trials. One third of the trials featured violations, the remainder featured grammatical variants in the language.

Participants in each condition saw the 16 variants they were trained on (with each object appearing along with half of the variants), plus 8 ungrammatical phrases (two Adj Only, two Num Only, and four two-modifier). Ungrammatical phrases included either word order or morphological violations. Word order violations consisted of pre-nominal modification in the morphology conditions, and the two unattested word orders in the word order conditions (i.e., Adj Num N and Num Adj N in Experiments 6 and 7 and Adj Num N and N Num Adj in Experiment 5). Across conditions, morphological violations consisted of unseen modifier suffixes. In the morphology conditions, all ungrammatical one-modifier phrases consisted of word order violations: they featured pre-nominal modification. In the word order conditions, ungrammatical one-modifier phrases consisted of morphological violations: modifiers featured different suffixes (e.g. *nedro* instead of *nefri*). Across conditions, half of the ungrammatical two-modifier phrases contained word order violations and the other half, morphological violations.

4.7.1.2 Post-experimental questionnaire

After participants completed the experiment and the grammaticality judgement task, they were asked to fill in a questionnaire (displayed on the monitor in a web browser). Participants were asked to estimate the relative frequencies of the majority variants in the input languages. They did so by selecting from a scale of 0 to 100% with 10% intervals.

4.7.2 Results

4.7.2.1 Grammaticality judgement

Morphology Participants in morphology conditions perfectly distinguished between grammatical and ungrammatical phrases: the proportion of correct judgements was very high for both Morphology in Experiment 5 (median = 1.00, MAD = 0.00) and Morphology Dyads in Experiment 7 (median = 0.958, MAD = 0.042).

Word order Scores in the word order conditions were also high across experiments (median = 0.917, MAD = 0.00). The majority of participants correctly judged phrases to be grammatical 91.7% of the time, which indicated two errors out of the 24 trials. An inspection of the pattern of errors revealed that participants in this condition failed to reject the word orders

absent from the training data as ungrammatical. These results suggest that participants assume word order is free in the input grammars.

Altogether these results suggest that whereas participant in the morphology conditions learn the restrictions of their input grammars, participants in the word order conditions do not, i.e., they perceive word order to be free.

4.7.2.2 Frequency reports

To determine whether encoding error contributed to participants' regularisation behaviour in production, participants were required to fill out a post-experimental questionnaire where they were asked to report estimates of the frequencies with which they read/heard majority input variants during training. If participants' reports of input frequencies differ significantly from their productions, this will suggest that the regularisation bias is in production rather than learning or encoding.

For each of the NP types, Figures 4.11 (isolate production) and 4.12 (communicative interaction) show the reported proportions of majority input variants in the input languages compared to the actual proportions in input and output languages. (For ease of comparison across experiments, I exclude the Word Order condition in Experiment 5 in this section.) I ran logistic mixed-effects regression models to explore the effect of encoding error on regularisation. In order to avoid problematic collinearity, I ran separate models for one-modifier NPs and for two-modifier NPs. However, model structures and data treatment were identical across. Training (input language) and production (output language) trials were dummy coded according to presence or absence of the majority input variants in each trial. In order to be able to compare production to reported relative frequencies in the questionnaire as the same dependent variable, I converted the input and reported relative frequencies into dummy coded absolute frequencies weighted accordingly to the number of input and output trials respectively. For example, if a participant reported 70% of the two-modifier majority variant N Adj Num in their testing and they produced two-modifier variants 30 times, I would code 0.7×30 presences and 0.3×30 absences of the majority input variant respectively. I entered four fixed effects into the model: NP Type (three ordered levels: Num Only and Adj Only or two-Mod for the one-modifier and two-modifier models respectively), Linguistic Level (morphology and word order), Response Type (three levels: Input Report as reference, Input Language and Output

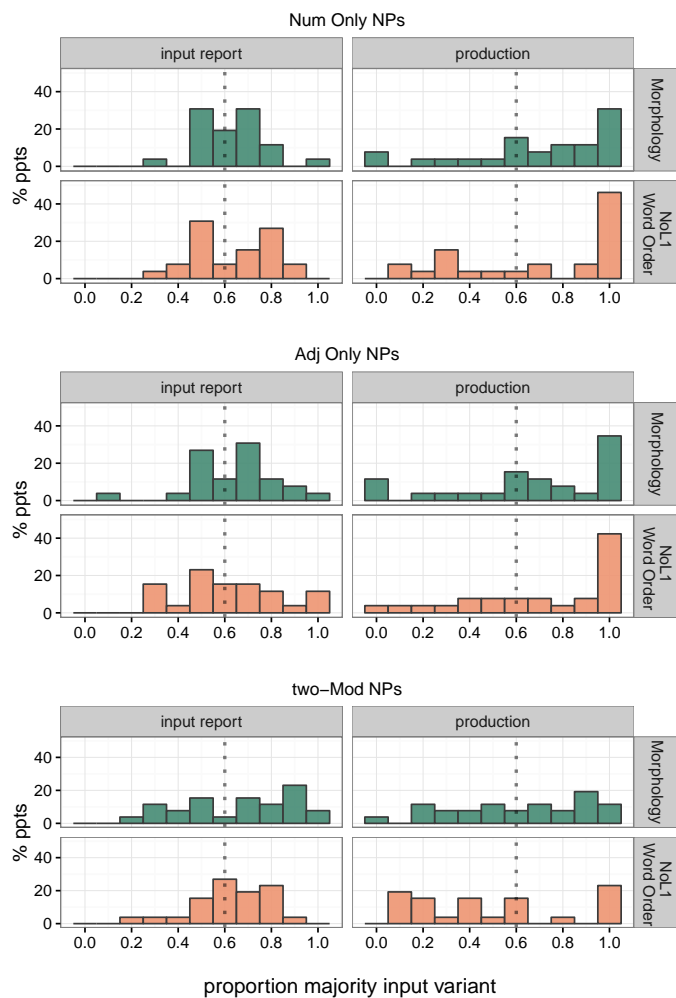


Figure 4.11: Proportions of majority input variants in Morphology (green) and NoL1 Word Order (orange), from Experiments 5 and 6 respectively (isolate production). From left to right, the estimate proportion of the majority input variant observed during training reported by participant in the post-experimental questionnaire (input report) and the actual proportion produced during testing. Results are shown for Morphology (green) and NoL1 Word Order (orange) conditions, for each NP type. Input proportions of the majority variants are signalled by a grey dashed vertical line.

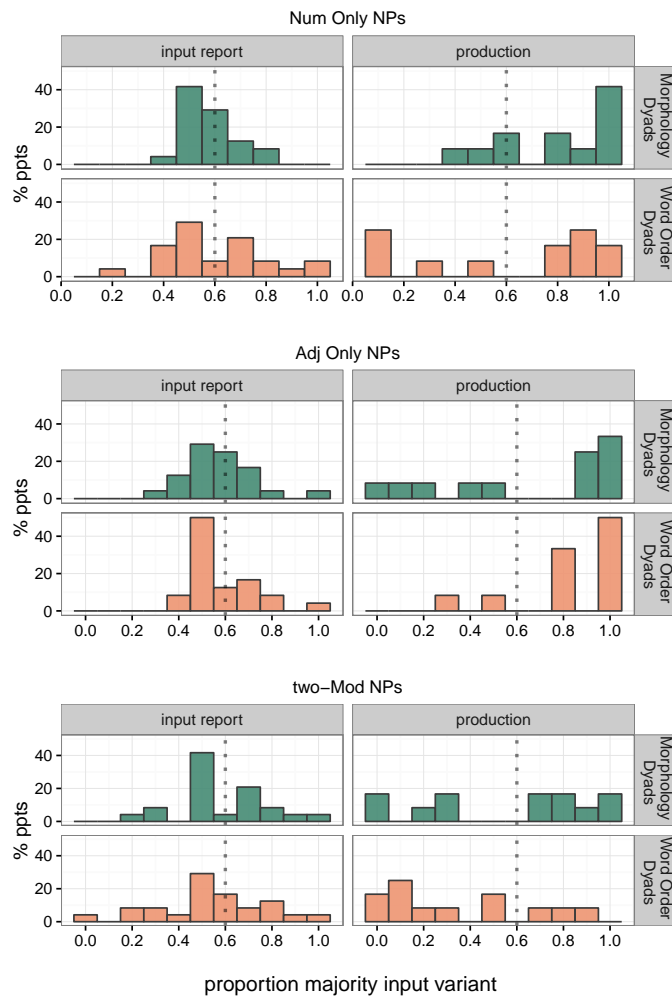


Figure 4.12: Proportions of majority input variants. From left to right, the estimate proportion of the majority input variant observed during training reported by participant in the post-experimental questionnaire (input report) and the actual proportion produced during testing. Results are shown for Morphology Dyads (green) and Word Order Dyads (orange) conditions, for each NP type. Input proportions of the majority variants are signalled by a grey dashed vertical line.

Language), and Experiment (Isolate production as reference, and Communicative interaction). I used reverse Helmert contrasts across fixed effects except for Response Type. I entered all interactions between fixed effects. As random effects I included random intercepts for Subject with Interlocutor (each participant in a dyad) nested as well as a by-Subject random slope for the effect of Response Type.

Results from the one-modifier model show that reported input frequencies did not differ significantly from actual input frequencies ($\beta = -0.053, SE = 0.086, p = 0.534$). On the other hand, production frequencies significantly differed from input reports ($\beta = 1.172, SE = 0.263, p < 0.001$). I also found that production of majority input variants differed from the reported input even more in Adj Only ($\beta = 0.417, SE = 0.0916, p < 0.001$), especially in Experiment 7 with the inclusion of communicative interaction ($\beta = 0.311, SE = 0.0916, p < 0.001$). These results suggest that the over-regularisation of majority variants in one-modifier phrases might be better explained by a production bias rather than by a learning bias or encoding error.

However, results from the two-modifier paint a somewhat more complex picture: participants' input reports do not differ significantly from the actual input frequencies ($\beta = 0.1033, SE = 0.117, p = 0.3792$) but neither does their production ($\beta = -0.314, SE = 0.283, p = 0.268$). A visual inspection of Figures 4.11 and 4.12 suggests that participants indeed do not over-produce majority variants in two-modifier phrases as they do in one-modifier phrases. No significant effects were found altogether (largest: $\beta = -0.202, SE = 0.108, p = 0.061$). Unlike with one-modifier phrases, I find that no clear conclusions can be drawn from this data provided the scarce over-regularisation of majority input variants and the limited data on participants' encoding. Unlike for one-modifier phrases, the input languages comprise four variants for two-modifier phrases and thus four different frequencies to track. Moreover, due to time constraints on attention span, post-experimental questionnaires were designed to only record encodings of majority input variants and thus only of one out of four different frequencies of two-modifier phrases. Thus considering that participants do indeed regularise (see section 4.6.2.2) as well as the the lack of over-production of two-mod variants, it is impossible to conclude whether regularisation of two-modifier phrases is driven by encoding or production biases with this data.

Altogether, results from the post-experimental questionnaires suggest a production bias for more regular systems rather than a learning or encoding bias for one-modifier phrases.

However, no conclusion can be drawn for two-modifier phrases.

4.7.2.3 Discussion: the roots of regularisation behaviour

Regularisation behaviour during individual learning and use has been proposed to be due to a bottleneck in memory which could affect the encoding of variants and their relative frequencies in training and/or in variant retrieval during production (Hudson Kam & Chang 2009). In this section, I investigated whether the regularisation found in *production* in the experiments was also found in encoding. I did so through grammaticality judgements, and post-experimental questionnaires in which participants are asked to estimate the relative frequencies of variants in the input. Results of the grammaticality judgement task suggest that participants had no problem in recognising grammatical variants in their input languages across conditions and experiments. Participants did indeed encode all variants, and not only the most frequent ones. I also found that the perception of variation differed between conditions. Participants in the word order conditions did not learn the restricted set of word order options, but instead perceived word order to be completely free. The consistency of these results across participants suggests that arbitrary restrictions in word order rules are very unlikely to be learned. By contrast, participants in the Morphology condition did not judge as grammatical morphological variants other than those they were exposed to during training. This difference in the perception of variation suggest that participants in word order conditions perceived a more variable system than participants in the morphology condition; nevertheless, the perceived additional complexity did not lead to significantly stronger regularisation behaviour (cf. Hudson Kam & Newport 2009; Perfors 2012a).

By contrast, results from the post-experimental questionnaire suggest that participants report the input frequencies of the majority input variant quite accurately. Thus although participants regularise their input in their productions, they nevertheless correctly estimate the input probabilities for one-modifier phrases. However, given that input frequencies of majority variants are 60%, it is difficult to discern between accuracy in encoding and perception of random variation (i.e., 50%). In any case, these results show that regularisation in production does not match participants' frequency encodings for one-modifier phrases. For two-modifier phrases, on the other hand, participants do not over-produce majority variants and their use in production matches the frequency encodings of the same variant in the input. From these results I

cannot conclude either limits in frequency encoding or a source of regularisation in production. I would require the estimates of all the different variants in two-modifier phrases to be able to test whether the source of regularisation stems from limits in frequency encoding.

In line with [Ferdinand et al. \(2017\)](#), these findings suggest that regularisation behaviour in production is not necessarily rooted in a memory bottleneck affecting the encoding of less frequent variants, at least for less variable phrases (i.e., one-modifier NPs), for which there were only two variables and frequencies to track. Nevertheless, regularisation behaviour could still be due to problems with lexical retrieval. Although participants can correctly recall input frequencies and the different variants in the input, they might struggle to retrieve them without being cued during production; [Hudson Kam and Chang \(2009\)](#) show that participants regularise their input less when lexical items are constantly available to them. Memory limitations can interact with mechanisms at play during production such as priming, which might impede the production of alternative variants to the ones immediately preceding and could affect either production in isolation and during communication. However, lexical retrieval could only hinder the production of variation in the morphology conditions, not in word order conditions because in the latter lexical items remain the same amongst variants. Altogether, these results lead us to conclude that regularisation also occurs (at least partly) during production.

4.8 General discussion

4.8.1 Regularisation behaviour is uniform across linguistic levels

In the present study I explored the extent to which level-specific biases interact with regularisation behaviour in language learning and use. Results showed that language users in fact regularise unconditioned variation in a similar way across morphology and word order, suggesting that a simplicity bias may be driven by a single, non-level-specific mechanism. Thus the present study does not provide supporting evidence for level-specific bias driving the asymmetries between regularisation of morphology and word order discussed in the language acquisition ([Anderssen et al. 2010](#); [Raymond et al. 2009](#); [Slobin 1966](#)) and pidgin/creole formation literature ([Bakker 2008](#); [Drechsel 1981](#); [Siegel 2006](#)). On the other hand, these results provide experimental evidence for the hitherto untested assumption that regularisation behaviour can be uniform across linguistic levels. However, this study constitutes only the first step towards

uncovering cross-level differences and similarities in the interaction between linguistic structure and learnability (or simplicity more broadly) pressures at play in language learning and use.

Comparable regularisation behaviour across levels in this study is conditioned on two factors. Firstly, the complexity of the systems of variation is also comparable across levels. Secondly, there are no specific biases targeting alternative variants available only at one level. If the first of the conditions is not met, according to the experimental evidence available (Hudson Kam & Newport 2009), we expect that regularisation behaviour will vary accordingly; the more complex the system of variation, the more it will be regularised. On the other hand, if the second of the conditions is not met, as we showed in Experiment 6, and alternative variants to the majority one are favoured or disfavoured, participants will accordingly over- or under-produce majority variants leading to overall more regular systems. Altogether, these results suggest that any asymmetry between levels in the context of L2 learners is more likely to be due to the inherent complexity of level-specific systems of variation; provided by the amount of variation within a given level or by its interaction with other features—within and/or between input languages.

4.8.2 Regularisation behaviour is comparable between isolate production and communicative interaction

Additionally, this study suggests that regularisation in contexts of L2 production cannot be reduced to the communicative context (cf. Klein & Perdue 1997; McWhorter 2001) and neither can the differences between linguistic levels (cf. Jansson, Parkvall, & Strimling 2015). I found that participants regularised to similar degrees in isolate production and during communicative interaction, suggesting that simplification of linguistic paradigms has to be also sought from learning biases or production biases at the individual level which are not induced by the communicative context.

My results contrast with previous studies (Fehér et al. 2017; Fehér, Wonnacott, & Smith 2016), which concluded that the communicative context triggers stronger regularisation behaviour. I proposed that differences between regularisation behaviour in the present study and Fehér, Wonnacott, and Smith (2016) are likely to be due to differences in the complexity of the input systems. Unlike in the present study, participants in Fehér, Wonnacott, and Smith (2016)

and Fehér et al. (2017) do not regularise the input variation in isolate production, i.e., they match the input probabilities on average instead. Studies demonstrating probability matching behaviour in adults often consider systems of two variants, which appear with different probabilities (e.g., Fehér, Wonnacott, & Smith 2016; Hudson Kam & Newport 2005; Realí & Griffiths 2009) or equiprobably (Experiment 2 in Fehér, Wonnacott, & Smith 2016). The experimental evidence available suggests that adult learners regularise more complex systems of variation (Hudson Kam & Newport 2009) and also that language users do not often eliminate all variation in the input during communicative interaction (Fehér, Wonnacott, & Smith 2016); these two results together, in combination with the relatively significant misalignment between pairs of participants in Experiment 7, explain why I do not find an effect of communicative interaction on regularisation behaviour in this study.

Other factors that could explain it are differences in task framing and/or in misalignment between participants in dyads. It is possible that Fehér, Wonnacott, and Smith (2016) emphasised the need for alignment between interlocutors more as they included a bonus for communicative accuracy; being rewarded for communicative accuracy might have shifted the focus away from faithfully reproducing the input language even more. However, it is worth noting that communicative accuracy was extremely high across studies, which at least rules out the possibility that effort to communicate effectively might lead to stronger regularisation. It is also possible that I did not find an effect of communicative interaction because participants within dyads did not align very well, potentially worse than in Fehér, Wonnacott, and Smith (2016) and Fehér et al. (2017). Misalignment could at the same time be caused by an anti-coordination bias whereby participants act cooperatively to reproduce the input variability more closely. However, I cannot provide any evidence for the effect of these factors. I thus conclude that differences between studies are most likely an artefact of the complexity of the input systems of variation.

4.8.3 Regularisation behaviour takes place in production as well as in learning

This study has demonstrated that regularisation takes place in the context of artificial language learning and that this cannot be reduced to the communicative context as it happens to similar degrees in isolate production and during communicative interaction. Moreover, results from grammaticality judgement tasks and post-experimental questionnaires also show that regularisation is a product of both learning and production biases (Ferdinand et al. 2017; Perfors &

Burns 2010), independent from the communicative context.

Across experiments and conditions, we observed that participants encoded all variants without problem, and not only the most frequent ones. Results from grammaticality judgements suggest that participants in the morphology conditions perfectly encoded all input variants and did not accept any that were not included in their training. Participants in the word order conditions were also able to identify input variants perfectly, however, they perceived higher variability in their input systems; instead of learning the restricted set of two-modifier phrases, they perceived word order to be free. These results suggest that regularisation behaviour was not due to error in the encoding of minority variants (Ferdinand et al. 2017; Perfors 2012b). Further evidence supporting regularisation in production comes from the results of the post-experimental questionnaires; participants report the input frequencies of one-modifier majority variants quite accurately. Thus although participants regularise the input in during production, they nevertheless correctly estimate the input probabilities for one-modifier phrases. Altogether these results follow Ferdinand et al. (2017) in showing that adult learners regularise unconditioned probabilistic variation in the input even when they are able to accurately report the input probabilities.

4.9 Conclusion

The literature on regularisation behaviour combines substantial data from language learning, transmission and use in the attempt to explain universal tendencies of human languages, providing the perfect platform to start exploring the effects of linguistic level and units in ubiquitous processes in language evolution. The present experimental paradigm (Culbertson et al. 2012; Hudson Kam & Chang 2009; Hudson Kam & Newport 2005) is ideally fit to explore the conditions under which language learners modify their input to produce languages that better conform to their biases and thus to investigate the link between these biases and typological generalisations. Nevertheless, in order to spell out asymmetries between linguistic levels in the structure and processes of natural languages, we require a better depiction of the different natures of the biases and their interaction with language types during learning and use. The present study has aimed to contribute to such need by providing evidence for commensurate regularisation strengths of probabilistic variation in morphology and word order during pro-

duction.

Language learners regularise complex systems of variation in their productions, suggesting a relationship between individual biases and processes of regularisation in natural languages. Nevertheless, the relationship between regularisation biases and asymmetries in regularisation processes between linguistic levels cannot be inferred from our results. Regularisation biases apply with similar strengths across linguistics levels given input languages with comparable initial complexity. Nevertheless, preferences for certain patterns within a linguistic level might in fact vary the strength of regularisation behaviour within the given level. Our study suggests that asymmetries in simplification processes in language formation ought to be sought from asymmetries in the input complexity of linguistic paradigms across levels and units and the overlap of traits across contributing languages.

Chapter 5

Summary and conclusions

5.1 Aims and contributions

In this thesis I have investigated the impact of language learning and use on absolute linguistic complexity, in particular on morphology and syntax. Through three sets of experimental studies, I explored the dynamics of linguistic complexity over various time-scales: its alteration during language learning and/or over the course of communicative interaction, and its evolution over cultural time. I built on previous work in the study of language change and evolution through cultural transmission (e.g Kirby et al. 2008, 2015; Verhoef 2012; Winters et al. 2015) to more precisely understand in which ways a drive towards efficiency in language learning and use (i.e., the reduction of relative complexity) might shape the complexity of linguistic systems and structures (i.e., absolute complexity) across the aforementioned different time-scales.

The majority of this thesis is devoted to a series of behavioural experiments that allow the direct investigation of causal relationships between relative and absolute complexity during language learning and use. Two basic steps were required to systematically investigate these relationships. The first one was to arrive at a tractable characterisation of complexity. Given the variety of approaches to measuring linguistic complexity, in Chapter 1 I proposed a taxonomy of absolute and relative complexities (see Figure 1.1). Broadly speaking, absolute complexity was defined in terms of the number of parts of a system or structure, and the directness of form-meaning mappings (i.e., the more direct, the simpler); for a reminder of the specific types of absolute complexity, see section 1.2.3. I argued that relative complexity, on the other hand, is best thought of in terms of effort in language learning and communication. Relative

simplicity is thus best thought of in terms of efficiency. Efficiency can be broadly defined as effort minimisation in unambiguously conveying meaning (Zipf 1949); an efficient system is one which maximises its learnability (i.e., generalisability) without compromising communicative effectiveness (Kirby et al. 2015).

The second step was to instrumentalise the learning and communicative contexts of language transmission in the laboratory. I drew upon already existing experimental techniques in artificial language learning (e.g. Culbertson et al. 2012; Hudson Kam & Chang 2009), interaction studies (e.g. Galantucci 2005; Krauss & Weinheimer 1964, 1966) and iterated learning (e.g. Kirby et al. 2008, 2015) that ultimately study the same question: how language learning and use impact language structure. These experimental methods have proven to be of great value for investigating language as a complex adaptive system shaped by the pressures acting on language users during learning and communication. This thesis has further shown that these same methods are also well suited to elucidate the relationship between relative and absolute complexity within a learner/user's individual behaviour and in the evolution of language as a system of behaviours shared at the population level.

5.1.1 Summary of experimental results

Languages are culturally transmitted through a repeated cycle of learning and communicative interaction. I departed from the assumption that these two aspects of cultural transmission impose interacting pressures that, along with neutral evolutionary processes, can shape the evolution of linguistic systems and their structure. I proposed that a drive towards the reduction of relative complexity in language learning and communication can shape absolute linguistic complexity during use and over cultural time. Table 5.1 provides a summary of how linguistic complexity is shaped by learning and communication in the light of the findings in this thesis. More specifically, it shows the linguistic contexts wherein participants were required to perform across studies (i.e., intergenerational transmission, learning and communication), and the changes in absolute complexity observed (see section 1.2.3). In the following paragraphs I will discuss more articulately how the relative complexity brought by the linguistic contexts impacts absolute complexity in the different experiments.

The experimental work presented in Chapter 2 looked at the cultural evolution of complex compositional structure from holistic (unstructured) languages. The findings in this chapter

		Chapter 2: Holistic to hierarchical compositional language	Chapter 3: Communicative context in L1	Chapter 4: Regularisation of variation
context	intergenerational transmission	✓		
	artificial language learning	✓		✓
	communicative interaction	✓	✓	✓
absolute complexity	system: <i>paradigmatic</i>	-	-	-
	<i>organisational</i>	-		-
	structural: <i>syntagmatic</i>	+	=	
	<i>hierarchial</i>	+	=	
	form-meaning <i>isomorphism</i>	+		
	mappings: <i>transparency</i>	+	+	+

Table 5.1: Evolution and modification of absolute complexity by study. I show the linguistic contexts wherein users were asked to perform in the experiments, and the reduction and increase of different absolute complexity types in the experimental results. Check-marks indicate the presence of the linguistic context. Plus and minus signs signify increase and reduction of a given type of absolute complexity respectively. Equal signs signal the lack of an effect. The absence of a sign indicates the lack of a test of this type of complexity in the study.

indicate that a complex meaning space paired with a learning bottleneck in transmission and a pressure for expressivity can result in compositional hierarchical constituent structure. Hierarchical compositional structure grants a *learnable* productive and productively interpretable language; it only requires learners to acquire a finite lexicon and a finite set of combinatorial rules (i.e., a grammar). The shift from unstructured systems to hierarchical compositional structure with regular combinatorial rules (see first column in Table 5.1) brings along an increase in structural complexity (syntagmatic and hierachical) and a reduction of system complexity (paradigmatic and organisational). Moreover, systems become isomorphic and their morphology transparent (i.e., mainly one-to-one form-meaning mappings). The evolution of structural complexity in the form of compositional structure is thus in the interest of the system's learnability (Culbertson & Kirby 2015; Mufwene 2013; K. Smith & Kirby 2012). Further support for system-optimisation is provided by the cultural evolution of regularity in word formation rules (e.g., plurality by suffixation only) and word order (e.g., fixed constituent order); this minimises the effort required to achieve productivity and productive interpretability and reduces the overall system complexity, both paradigmatic (i.e., fewer and more productive forms) and

organisational (i.e., no word order variation).

Chapter 2 also investigated the precise role of communicative interaction in combination with a learning bottleneck. Previous work has mainly characterised communicative interaction as a natural promoter of expressivity in cultural transmission (for an exception, cf. [Winters 2017](#)). I contrasted an artificial pressure against ambiguity with communicative interaction. In comparison to the artificial pressure, communicative interaction facilitated the evolution of systematicity and structure. I concluded that coordination in communicative interaction plays an important role in language evolution. During communication, users need to coordinate on learnable as well as expressive conventions, to solve not only the immediate task at hand but also future interactions. This trade-off between learnability and expressivity in coordination—in combination with language transmission—speeds up the emergence of efficient and structured linguistic systems over repeated usage. I provided further support for the importance of coordination in language evolution by demonstrating that if coordination processes between interlocutors are hindered (e.g., the piecemeal fashion in which conventions are established), so is the evolution of linguistic structure. I thus conclude that the evolution of complex compositional structure and regularity in combinatorial rules is best characterised as a product of the trade-off between at least three pressures at play in cultural transmission: a pressure for coordination, a pressure for learnability and a pressure for expressivity (see also [Winters 2017](#)).

I further addressed the unique effect of communicative interaction on linguistic complexity in Chapter 3 by removing (artificial) language learning and transmission completely. Speakers used their native language to express novel meanings either in isolation or during communication. I demonstrated that the communicative context—where feedback is provided—leads to more efficient and simpler linguistic systems: interlocutors output more productive and transparent morphological lexicons (i.e., lower paradigmatic complexity and clearer form-meaning mappings). There is thus a drive for simplicity imposed by the communicative context that does not show up in isolated production with equally expressive systems. I proposed that this difference could also be explained by coordination and its interaction with learnability pressures during communication not specifically related to language learning but to generalisability more broadly; in order to maximise the efficiency of a system of conventions, users need to be conservative towards previously successful solutions to coordination problems and be able to generalise them to new data. Taken together, this minimises variability in the lexicon.

These first two chapters provided support for the claim that morphological and syntactic complexity is shaped by an overarching drive towards simplicity to maximise learnability without jeopardising communicative effectiveness. The experiments in Chapter 4 assessed the uniformity of this simplicity bias across different linguistic levels which at the same time pertain to different types of system complexity. In this set of experiments, I built on previous work combining statistical learning and artificial language learning techniques (Culbertson et al. 2012; Fehér, Wonnacott, & Smith 2016; Hudson Kam & Newport 2005) to compare regularisation of unconditioned variation across morphology and word order separately. Variation in morphology and word order are quantitatively comparable but represent two aspects of system complexity, i.e., paradigmatic and organisational complexity respectively. I showed that language users regularise unconditioned variation to a similar degree across linguistic levels, suggesting that the simplicity bias may be driven by a single, non-level-specific mechanism. Regularisation results in a reduction of system complexity; paradigmatic in the case of morphology, and organisational in the case of word order. By reducing system complexity, form-meaning mappings become more transparent, which increases both efficiency and effectiveness in communication; one-to-one mappings facilitate retrieval in production and comprehension, and help contrast new and old information. Moreover, I compared regularisation across levels in communication and in isolate production. In line with the aforementioned drive towards efficiency triggered by the communicative context, previous work (Fehér, Wonnacott, & Smith 2016) had shown that language users regularise unconditioned variation more during communication than in isolate production. The results in Chapter 4 are not consistent with these findings: language users in fact regularised unconditioned variation in a similar way in the absence and in the presence of a communicative context. I argued that these results are most likely due to the difference in the complexity of the system of variation. Learners tend to probability match simple (i.e., two variants) systems of variation (see Fehér, Wonnacott, & Smith 2016). However, with more complex systems of variation, regularisation behaviour in isolate production increases (Hudson Kam & Newport 2009). Thus with more complex systems of variation, differences between isolate production and production during communication might decrease.

Chapter 4 also investigated the cognitive roots of regularisation behaviour. Regularisation in language learning is often assumed to be due to a memory bottleneck that impedes

the encoding of less frequent variants and their frequencies (Hudson Kam & Newport 2009). Post-experimental questionnaires indicated that language users produce less variation than they estimate in the input. These results suggest that regularisation behaviour is not only due to problems in encoding; there is a drive towards regularisation acting on production as well (see also Ferdinand et al. 2017). I will come back to this in the next section.

Across these different time-scales, results are consistent with an overarching simplicity bias in language learning and use. Learners/users alter linguistic systems to maximise their learnability and communicative efficiency. Such a drive towards reducing the effort of unambiguously conveying a message generally surfaces as a reduction of system complexity and the establishment of transparent one-to-one form-meaning mappings. We observe in Table 5.1 that a reduction in system complexity—in particular paradigmatic complexity—and an increase in transparency are common processes in linguistic change during language learning and/or use in communication, and over cultural time. These effects of an overarching simplicity bias overlap with the general principles found across the most explicit (already existing) criteria for complexity both in absolute approaches (McWhorter 2001; Miestamo 2006a) and relative approaches (Kusters 2003); a general principle of economy—fewer parts and/or rules—and a general principle of transparency in form-meaning mappings. Altogether, these two principles maximise the generalisability of a system and its interpretability.

5.1.2 Roots of an overarching simplicity bias

A simplicity bias has been proposed as a unifying principle of learning within cognitive science (Chater, Clark, Goldsmith, & Perfors 2015; Chater & Vitányi 2003; Culbertson & Kirby 2015; Ferrer i Cancho et al. 2013). Much of learning involves finding patterns in data, and learners are biased towards simple patterns out of infinitely many patterns compatible with the data (Chater & Vitányi 2003). Specifically in the context of language learning, Culbertson et al. (2012) argue that “whatever other biases learners have when they face some learning problem, they are also likely to be applying an overarching simplicity bias”. What can we say about the roots of such an overarching simplicity bias in the light of the experiments in this thesis?

A simplicity bias is best thought of as a learnability bias (Culbertson & Kirby 2015), defined as a preference for generalisable behaviour. Learnability pressures imposed on the learner/user during language learning and communication might trigger the effects of a learn-

ability bias; over cultural time, the product of this bias accumulates, potentially leading to languages that are better fit for human learning. In Chapter 1 I discussed two different types of learnability pressures: a learning bottleneck provided by a limited amount of data and a memory bottleneck (a contrast already made in [Cornish 2010](#)). The findings in this thesis support these two types of bottlenecks in language learning. In Experiment 1 (section 2.2, Chapter 2) I showed that the combination of a learning bottleneck in transmission and a pressure for expressivity during production drive the evolution of linguistic structure in the absence of communicative interaction. In this experiment, both types of bottlenecks are present. Learners were trained on half of the data they were later tested on, which imposes a data bottleneck that triggers generalisability in production. In addition, learners cannot learn the 44 complex holistic form-meaning mappings they are exposed to during training (see section 2.6.1); this leads to a memory bottleneck.

Chapter 4 provides further evidence for the effect of a memory bottleneck in individual learning, this time in the absence of a data bottleneck. Experiments 5 and 6 suggest that learners cannot track the input frequencies of subsystems of variation of more than two variants (i.e., two-modifier phrases). However, they can track the input frequencies of less complex subsystems of variation (i.e., those of one-modifier variants). Altogether, these results suggest that although regularisation behaviour is partially driven by problems with encoding, it also occurs during production and might be even orthogonal to learning in some cases. These results are not surprising if we think of language as an effective system of communication ([Zipf 1949](#)). Unconditioned variation is redundant in a linguistic system and could slow down processing and hinder production. In order to reproduce probabilistic unconditioned variation and not regularise, users would need to refrain from the effects of priming and would need to keep in memory more items always available for retrieval (see [Hudson Kam & Chang 2009](#)).

The previous paragraphs have discussed different types of learnability pressures in individual language learning and production. Nevertheless, as posited throughout this thesis, learnability pressures are also present during communicative interaction. Over the course of communication, users coordinate with each other to establish linguistic conventions in order to communicate effectively and efficiently. Coordination requires users to pay attention to their interlocutor's behaviour so they can match it and correctly interpret meanings; moreover, linguistic conventions need to be generalisable to solve not only present but also future interaction

events. Only by highlighting the role of learning in coordination can we straightforwardly explain the differences between the experiments in Chapter 2 and those between conditions in Chapter 3. Experiment 2 in Chapter 2 suggests that the communicative context speeds up the evolution of systematic and structured languages (see also Carr et al. 2016), accompanied by an early drop in paradigmatic complexity and an early establishment of transparent form-meaning mappings (see section 2.3). Experiment 4 in Chapter 3 further indicates that communicative interaction leads to more efficient linguistic systems, with simpler morphological lexicons and greater transparency. However, results from Experiment 7 in Chapter 4 imply that learners regularise complex systems of unconditioned variation to similar degrees during isolated production and during communicative interaction; post-experimental questionnaires suggest that they encode input frequencies in similar ways as well. I proposed that in the presence of stronger simplification biases in individual learning and production, the addition of learnability pressures during communication does not make a difference; however, there is no experimental evidence yet available to further back up this claim. Further work is required to pin down the different factors that trigger generalisation during communication (e.g., memory vs. data bottlenecks) and the role of interaction in simplification processes more generally.

In sum, in the light of the work presented in this thesis, a simplicity or learnability bias is rooted in memory and data bottlenecks in language learning and transmission (see also Cornish 2010; Culbertson & Kirby 2015), in processing constraints in production (see also Hudson Kam & Chang 2009), and in coordination during communicative interaction (see also Winters 2017). However, further work is needed to characterise simplicity pressures and biases in production and communication.

5.2 Directions for future research

This thesis has demonstrated that laboratory experiments using artificial language learning are an effective methodology to provide direct behavioural evidence of individual biases in language learning and use, and their impact on linguistic systems and structure over time. The experimental work presented constitutes one of the first attempts to apply these methods to explore the relationship between relative and absolute complexity in language evolution (e.g., see Atkinson 2016; Bentz & Berdicevskis 2016; Fehér et al. 2017; Fehér, Wonnacott, & Smith

2016; Tinitis et al. 2017). And to my knowledge, it is the first attempt to explore how uniform a simplification bias is across learning and communication, and across different linguistic units.

There are a number of directions for future research stemming from this thesis, related directly to the experiments or to the study of the cultural evolution of linguistic complexity more broadly. It is clear from the experimental evidence discussed that the link between learners'/users' biases and language structure is complex and the roots of these biases are far from clear. Future studies should help tease apart the different types of learnability pressures found in language learning and use, i.e., help characterise different aspects of the posited overarching simplicity bias (Chater & Vitányi 2003; Culbertson & Kirby 2015). Most of the work presented in this thesis comes from a tradition that started from the perspective of individual learning (e.g. Brighton et al. 2005; Kirby et al. 2008; Kirby & Hurford 2002) and only recently has been incorporating communication into the models (e.g. Fehér et al. 2017; Fehér, Wonnacott, & Smith 2016; Kirby et al. 2015; Winters et al. 2015). As a consequence, the different types of learnability pressures at play in communication and their role in language evolution are not as well characterised as they are in language learning and transmission. Further work needs to systematically compare learning mechanisms in coordination and individual language learning. We could start by contrasting memory and data bottlenecks in communication and comparing their respective effects with those of individual language learning. We should also systematically contrast encoding and production. In artificial language learning studies, production is usually taken to be a reflection of what learners encode during learning. However, as suggested in Chapter 4, users might be biased to simplify the learned input during production, irrespective of how well they learned it. There might be types of production constraints—some in memory (Hudson Kam & Chang 2009)—which lead to similar behavioural products to those of encoding errors. If we find similar simplification patterns in production and in encoding (as suggested in Chapter 4 here and in Ferdinand et al. 2017), it would suggest that cognitive biases at play during early language learning might turn into latent preferences later on even with higher proficiency. Clarifying the connection between production and learning biases will help strengthen cultural transmission as a solution to the linkage problem (Kirby 1999)—i.e., how individual behaviours penetrate linguistic systems. Finding similar behavioural products in language learning and use could help solve the problem posited by the fact that language is transmitted with higher fidelity with more exposure during development or over time. Regard-

less of linguistic proficiency, the behavioural product would still be shaped by an overarching simplicity (or learnability) bias than can equally penetrate the system over cultural time.

Another direction for future research is to design the same behavioural experiments for cross-linguistic comparison. In the framework adopted in this thesis, I make inferences with regard to a *general* learner/user. However, as experimentalists, we need to take into account that prior linguistic knowledge affects participants' behaviour (as we saw in section 4.4, Chapter 4). Running the same experiments with different linguistic populations would help us discern universal from language-specific patterns linking relative complexity to absolute complexity in language evolution. Additionally, corpus studies can also be used to test the hypotheses extracted from the results in behavioural experiments cross-linguistically. For instance, the work presented in Chapter 4 could be complemented with a corpus analysis of input complexity—measured comparably—of contact languages across linguistic units, and the difference with output complexity (for a first attempt, see Good 2015).

Lastly, and crucially, future work has to explore the emergence of grammatical categories, and of system complexity more broadly. In this thesis, we explored complexification and simplification processes driven by selective evolutionary pressures in cultural transmission—learnability, expressivity and coordination. However, the emergence of complexity has only been shown in Chapter 2 at the level of structure—hierarchical and syntagmatic complexity—and in the interest of system simplicity. System complexity, on the other hand, has always been the subject of simplification across experiments. Moreover, Chapter 2 tests predictions about the evolution of lexical morphology—corresponding to features of the meaning—and not about the evolution of purely grammatical morphology—without direct correspondence to any meaning feature. We need to explore the mechanisms by which system complexity emerges or increases and in particular how purely grammatical categories emerge. This line of research would at the same time further highlight the need to study the relationship between different aspects of complexity in a system. Previous work has shown that complexity in certain linguistic domains can affect complexity in others; e.g. the emergence of case marking might emerge so as to increase communicative effectiveness in a system with highly variable word order (see Fedzechkina, Jaeger, & Newport 2011; Futrell, Mahowald, & Gibson 2015; Montemurro & Zanette 2011).

5.3 General conclusion

The experimental evidence presented in this thesis supports the hypothesis that the cultural and cognitive pressures acting on language users to minimise relative complexity during learning and communicative interaction—for learnability, expressivity and coordination—are at least partially responsible for the evolution of absolute linguistic complexity. In particular, a drive towards efficiency in learning and communication leads to the simplification of system complexity—across linguistic levels—and to the establishment of transparent form-meaning mappings; both together help minimise the effort of unambiguously conveying a message. Structural complexity can also emerge in the interest of system simplicity and productivity. I thus concluded that a bias towards simplicity is a crucial driver of language change and evolution more broadly.

The approach taken in this thesis promotes a view of linguistic complexity as an evolving variable—and languages as complex adaptive systems more broadly—determined by the biases of learners and users as languages are culturally transmitted. Altogether, I believe the present work has shown that linguistic complexity, can indeed be a “mask for simplicity” (Simon 1996), that is, a simplicity determined by human cognition and culture in language evolution.

References

- Akhtar, N., & Gernsbacher, M. A. (2007). Joint attention and vocabulary development: a critical look. *Language and linguistics compass*, 1(3), 195–207. doi: [10.1111/j.1749-818X.2007.00014.x](https://doi.org/10.1111/j.1749-818X.2007.00014.x)
- Anderssen, M., Bentzen, K., & Westergaard, M. (2010). *Variation in the input: Studies in the acquisition of word order* (Vol. 39). Springer. doi: [10.1007/978-90-481-9207-6](https://doi.org/10.1007/978-90-481-9207-6)
- Ansaldò, U., Matthews, S., & Lim, L. (2007). *Deconstructing creole* (Vol. 73). Amsterdam/Philadelphia: John Benjamins.
- Atkinson, M. (2016). *Sociocultural determination of linguistic complexity* (Doctoral dissertation, The University of Edinburgh). Retrieved from <https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxtZGF0a2luc29uMDAwfGd4OjRkNzlmNzAzM2RmZTZhZWQ>
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Baggio, G., van Lambalgen, M., & Hagoort, P. (2012). The processing consequences of compositionality. *The Oxford Handbook of Compositionality*. Oxford, 657–674. doi: [10.1093/oxfordhb/9780199541072.013.0032](https://doi.org/10.1093/oxfordhb/9780199541072.013.0032)
- Bakker, P. (2008). Pidgins versus creoles and pidgincreoles. In (pp. 48–73). Oxford: Wiley-Blackwell. doi: [10.1002/9781444305982.ch6](https://doi.org/10.1002/9781444305982.ch6)
- Baronchelli, A., Loreto, V., & Steels, L. (2008). In-depth analysis of the naming game dynamics: the homogeneous mixing case. *International Journal of Modern Physics C*, 19(05), 785–812. doi: [10.1142/S0129183108012522](https://doi.org/10.1142/S0129183108012522)
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi: [10.18637/jss.v067.i01](https://doi.org/10.18637/jss.v067.i01)

- Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., ... Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language learning*, 59(s1), 1–26. doi: [10.1111/j.1467-9922.2009.00533.x](https://doi.org/10.1111/j.1467-9922.2009.00533.x)
- Beckner, C., Pierrehumbert, J. B., & Hay, J. (2017). The emergence of linguistic structure in an online iterated learning task. *Journal of Language Evolution*, lzx001. doi: [10.1093/jole/lzx001](https://doi.org/10.1093/jole/lzx001)
- Beckner, C., & Wedel, A. (2009). The roles of acquisition and usage in morphological change. In *Annual meeting of the berkeley linguistics society* (Vol. 35, pp. 1–12). doi: [0.3765/bls.v35i1.3593](https://doi.org/0.3765/bls.v35i1.3593)
- Bengfort, B. (2014). *Python-enabled Berkeley Parser*. <https://github.com/emcnany/berkeleyinterface>.
- Bentz, C., & Berdicevskis, A. (2016). Learning pressures reduce morphological complexity: Linking corpus, computational and experimental evidence. In *Coling 2016* (pp. 222–232). Retrieved from <http://www.aclweb.org/anthology/W16-4125>
- Bentz, C., & Winter, B. (2013). Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change*, 3(1), 1–27. Retrieved from http://www.bodowinter.com/papers/bentz_winter_L2_case.pdf doi: [10.1163/22105832-13030105](https://doi.org/10.1163/22105832-13030105)
- Bichakjian, B. H. (1988). *Evolution in language* (Vol. 18). Karoma Pub.
- Bickerton, D. (2015). *Roots of language*. Berlin: Language Science Press. doi: [10.17169/langsci.b91.109](https://doi.org/10.17169/langsci.b91.109)
- Bisang, W. (2009). On the evolution of complexity: sometimes less is more in east and mainland southeast asia. In G. Sampson, D. Gil, & P. Trudgill (Eds.), *Language complexity as an evolving variable* (Vol. 13). Oxford: Oxford University Press.
- Boulton, R. (2013). *Pystemmer 1.3.0 – snowball stemming algorithms for information retrieval*. <http://snowball.tartarus.org/>.
- Boyd, R., & Richerson, P. J. (1988). *Culture and the evolutionary process*. University of Chicago Press.
- Branigan, H. P., Catchpole, C. M., & Pickering, M. J. (2011). What makes dialogues easy to understand? *Language and Cognitive Processes*, 26(10), 1667–1686. doi: [10.1080/01690965.2010.524765](https://doi.org/10.1080/01690965.2010.524765)

- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6), 1482–93. doi: [10.1037//0278-7393.22.6.1482](https://doi.org/10.1037//0278-7393.22.6.1482)
- Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial life*, 8(1), 25–54.
- Brighton, H., Smith, K., & Kirby, S. (2005). Language as an evolutionary system. *Physics of Life Reviews*, 2(3), 177–226. doi: [10.1016/j.plrev.2005.06.001](https://doi.org/10.1016/j.plrev.2005.06.001)
- Brochhagen, T., Franke, M., & van Rooij, R. (2016). Learning biases may prevent lexicalization of pragmatic inferences: a case study combining iterated (bayesian) learning and functional selection. In *Proceedings of the 38th annual conference of the cognitive science society* (pp. 2081–2086). Austin, TX: Cognitive Science Society.
- Bruner, J., & Watson, R. (1983). *Child's talk: Learning to use language*. Oxford: Oxford University Press.
- Bybee, J. (2009). Grammaticization: Implications for a theory of language. In J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura, & S. Ozcaliskan (Eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of dan isaac slobin*. Psychology Press.
- Bybee, J., & Hopper, P. J. (2001). *Frequency and the emergence of linguistic structure* (Vol. 45). Amsterdam/Philadelphia: John Benjamins. doi: [10.1075/tsl.45](https://doi.org/10.1075/tsl.45)
- Bybee, J., & Slobin, D. I. (1982). Why small children cannot change language on their own: Suggestions from the english past tense. In A. Alqvist (Ed.), *Papers from the fifth international conference on historical linguistics* (pp. 29–37). Amsterdam: John Benjamins.
- Cangelosi, A., & Parisi, D. (2012). *Simulating the evolution of language*. Springer Science & Business Media.
- Carr, J. W., Smith, K., Cornish, H., & Kirby, S. (2016). The cultural evolution of structured languages in an open-ended, continuous world. *Cognitive science*, 41(4), 892–923. doi: [10.1111/cogs.12371](https://doi.org/10.1111/cogs.12371)
- Chater, N., Clark, A., Goldsmith, J. A., & Perfors, A. (2015). *Empiricism and language learnability*. Oxford: Oxford University Press. doi: [10.1093/acprof:oso/9780198734260.001.0001](https://doi.org/10.1093/acprof:oso/9780198734260.001.0001)

- Chater, N., & Vitányi, P. (2003). Simplicity: a unifying principle in cognitive science? *Trends in cognitive sciences*, 7(1), 19–22. doi: [10.1016/S1364-6613\(02\)00005-0](https://doi.org/10.1016/S1364-6613(02)00005-0)
- Chomsky, N. (1957). *Syntactic structures*. Berlin: De Gruyter.
- Chomsky, N. (1965). *Aspects of theory of syntax*. Cambridge, MA: The MIT Press.
- Chomsky, N. (1980). Rules and representations. *Behavioral and brain sciences*, 3(01), 1–15. doi: [10.1017/S0140525X00001515](https://doi.org/10.1017/S0140525X00001515)Published
- Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Chicago, IL: Greenwood Publishing Group.
- Christiansen, M. H. (2000). Using artificial language learning to study language evolution: Exploring the emergence of word order universals the evolution of language. In L. Ghadakpour & J. Dessalles (Eds.), *The evolution of language: 3rd international conference* (pp. 45–48). Retrieved from file:///Users/pplsuser/Downloads/Using_artificial_language_learning_to_study_langua.pdf
- Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and brain sciences*, 31(5), 489–509. doi: [10.1017/S0140525X08004998](https://doi.org/10.1017/S0140525X08004998)
- Claidière, N., & Sperber, D. (2007). The role of attraction in cultural evolution. *Journal of Cognition and Culture*, 7(1), 89–111. doi: [10.1163/156853707X171829](https://doi.org/10.1163/156853707X171829)
- Clark, E. V. (1988). On the logic of contrast. *Journal of Child Language*, 15(02), 317–335. doi: [10.1017/S0305000900012393](https://doi.org/10.1017/S0305000900012393)
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge university press.
- Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. Joshi, B. L. Webber, & I. Sag (Eds.), *Elements of discourse understanding*. Cambridge: Cambridge University Press.
- Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive science*, 13(2), 259–294. doi: [10.1207/s15516709cog1302_7](https://doi.org/10.1207/s15516709cog1302_7)
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22(1), 1–39. doi: [10.1016/0010-0277\(86\)90010-7](https://doi.org/10.1016/0010-0277(86)90010-7)
- Collier, K., Bickel, B., van Schaik, C. P., Manser, M. B., & Townsend, S. W. (2014). Language evolution: syntax before phonology? *Proceedings of the Royal Society of London B: Biological Sciences*, 281(1788), 20140263. doi: [10.1098/rspb.2014.0263](https://doi.org/10.1098/rspb.2014.0263)
- Cook, V. (1988). Designing a basic parser for call. *Calico Journal*, 50–67. doi:

[10.1558/cj.v6i1.50-67](https://doi.org/10.1558/cj.v6i1.50-67)

- Cornish, H. (2010). Investigating how cultural transmission leads to the appearance of design without a designer in human communication systems. *Interaction Studies*, 11(1), 112–137. doi: [10.1075/is.11.1.02cor](https://doi.org/10.1075/is.11.1.02cor)
- Crain, S., Goro, T., & Thornton, R. (2006). Language acquisition is language change. *Journal of Psycholinguistic Research*, 35(1), 31–49. doi: [10.1007/s10936-005-9002-7](https://doi.org/10.1007/s10936-005-9002-7)
- Croft, W. A. (2004). Form, meaning and speakers in the evolution of language: Commentary on kirby, smith and brighton. *Studies in Language. International Journal sponsored by the Foundation Foundations of Language*, 28(3), 608–611. doi: [10.1075/sl.28.3.10cro](https://doi.org/10.1075/sl.28.3.10cro)
- Culbertson, J., & Kirby, S. (2015). Simplicity and specificity in language: Domain-general biases have domain-specific effects. *Frontiers in psychology*, 6. doi: [10.3389/fpsyg.2015.01964](https://doi.org/10.3389/fpsyg.2015.01964)
- Culbertson, J., & Kirby, S. (2016). Simplicity and specificity in language: Domain-general biases have domain-specific effects. *Frontiers in Psychology*, 6, 1964. doi: [10.3389/fpsyg.2015.01964](https://doi.org/10.3389/fpsyg.2015.01964)
- Culbertson, J., & Newport, E. L. (2015). Harmonic biases in child learners: In support of language universals. *Cognition*, 139, 71–82. doi: [10.1016/j.cognition.2015.02.007](https://doi.org/10.1016/j.cognition.2015.02.007)
- Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, 122(3), 306–329. doi: [10.1016/j.cognition.2011.10.017](https://doi.org/10.1016/j.cognition.2011.10.017)
- Dahl, Ö. (2004). *The growth and maintenance of linguistic complexity* (Vol. 71). Amsterdam/Philadelphia: John Benjamins.
- Dahl, Ö. (2008). Testing the assumption of complexity invariance: the case of elfdalian and swedish. In M. Miestamo, S. Kaius, & F. Karlsson (Eds.), *Language complexity: Typology, contact, change* (pp. 3–22). Amsterdam: John Benjamins.
- Dahl, Ö. (2011). Grammaticalization and linguistic complexity. In B. Heine & H. Narrog (Eds.), *The oxford handbook of grammaticalization*. Oxford: Oxford University Press.
- Damerau, F. J. (1964). A technique for computer detection and correction of spelling errors. *Communications of the ACM*, 7(3), 171–176. doi: [10.1145/363958.363994](https://doi.org/10.1145/363958.363994)
- Davies, B. L. (2006). Testing dialogue principles in task-oriented dialogues: An exploration of cooperation, collaboration, effort and risk. *Leeds Working Papers in Linguistics and Phonetics*, 11, 30–64. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/>

- [download?doi=10.1.1.482.6216&rep=rep1&type=pdf](#)
- DeGraff, M. (1999). *Language creation and language change*. Cambridge, MA: The MIT Press.
- Deutscher, G. (2009). “overall complexity”: a wild goose chase? In G. Sampson, D. Gil, & P. Trudgill (Eds.), *Language complexity as an evolving variable* (Vol. 13, pp. 64–79). Oxford: Oxford University Press.
- Dhillon, R. (2010). Examining the ‘noun bias’: A structural approach. *University of Pennsylvania Working Papers in Linguistics*, 16(1), 7.
- Dietrich, R., Klein, W., & Noyau, C. (1995). The acquisition of temporality in a second language. , 7.
- Dimitriadis, A., Windhouwer, M., Saulwick, A., Goedemans, R., Bíró, T., et al. (2009). How to integrate databases without starting a typology war: The typological database system. In S. Musgrave, M. Everaert, & A. Dimitriadis (Eds.), *The use of databases in cross-linguistic studies* (pp. 155–207). Berlin: De Gruyter.
- Drechsel, E. J. (1981). A preliminary sociolinguistic comparison of four indigenous pidgin languages of north america (with notes towards a sociolinguistic typology in american indian linguistics). *Anthropological Linguistics*, 93–112. Retrieved from <http://www.jstor.org/stable/30027542>
- Dryer, M. S. (2013a). Coding of nominal plurality. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/33>
- Dryer, M. S. (2013b). Order of subject, object and verb. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/81>
- Dryer, M. S., & Haspelmath, M. (Eds.). (2013). *Wals online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/>
- Edmonds, B. M. (1999). *Syntactic measures of complexity* (Unpublished doctoral dissertation). University of Manchester.
- Ellis, N. C. (2013). Second language acquisition. In G. Trousdale & T. Hoffmann (Eds.), *The oxford handbook of construction grammar* (pp. 365–378). Oxford: Oxford University Press. doi: [10.1093/oxfordhb/9780195396683.001.0001](https://doi.org/10.1093/oxfordhb/9780195396683.001.0001)

- Ellis, N. C., & Larsen-Freeman, D. (2009). *Language as a complex adaptive system* (Vol. 3). Malden, MA: Wiley–Blackwell.
- Endres, D. M., & Schindelin, J. E. (2003). A new metric for probability distributions. *IEEE Transactions on Information theory*, 49(7), 1858–1860. doi: [10.1109/TIT.2003.813506](https://doi.org/10.1109/TIT.2003.813506)
- Engesser, S., Ridley, A. R., & Townsend, S. W. (2016). Meaningful call combinations and compositional processing in the southern pied babbler. *Proceedings of the National Academy of Sciences*, 201600970. doi: [10.1073/pnas.1600970113](https://doi.org/10.1073/pnas.1600970113)
- Esper, E. A. (1925). A technique for the experiment investigation of associative interference in artificial linguistic material. *Language monographs*.
- Esper, E. A. (1966). Social transmission of an artificial language. *Language*, 42(3), 575–580. doi: [10.2307/411408](https://doi.org/10.2307/411408)
- Farrar, M. J. (1992). Negative evidence and grammatical morpheme acquisition. *Developmental psychology*, 28(1), 90–98. doi: [10.1037/0012-1649.28.1.90](https://doi.org/10.1037/0012-1649.28.1.90)
- Fay, N., & Ellison, T. M. (2013). The cultural evolution of human communication systems in different sized populations: usability trumps learnability. *PloS one*, 8(8), e71781. doi: [10.1371/journal.pone.0071781](https://doi.org/10.1371/journal.pone.0071781)
- Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2011). Functional biases in language learning: Evidence from word order and case-marking interaction. In *Proceedings of the cognitive science society* (Vol. 33). Retrieved from <http://www.sas.upenn.edu/~mfedze/docs/FedzechkinaJaegerNewport11.pdf>
- Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences*, 109(44), 17897–17902. doi: [10.1073/pnas.1215776109](https://doi.org/10.1073/pnas.1215776109)
- Fehér, O., Ritt, N., & Smith, K. (2017). Asymmetric accommodation during interaction leads to the regularisation of linguistic variants. doi: [10.17605/OSF.IO/YBNQ9](https://doi.org/10.17605/OSF.IO/YBNQ9)
- Fehér, O., Ritt, N., Wonnacott, E., & Smith, K. (2016). Communicative interaction leads to the elimination of unpredictable variation. In S. Roberts, C. Cuskley, L. McCrohon, L. Barcelo-Coblijn, O. Fehér, & T. Verhoef (Eds.), *The evolution of language: Proceedings of the 11th international conference*. Retrieved from <http://evolang.org/neworleans/papers/137.html>

- Fehér, O., Wonnacott, E., & Smith, K. (2016). Structural priming in artificial languages and the regularisation of unpredictable variation. *Journal of Memory and Language*, *91*, 158–180. doi: [10.1016/j.jml.2016.06.002](https://doi.org/10.1016/j.jml.2016.06.002)
- Ferdinand, V., Kirby, S., & Smith, K. (2017). The cognitive roots of regularization in language. *arXiv preprint*. Retrieved from [arXiv:1703.03442](https://arxiv.org/abs/1703.03442)
- Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, *30*(2), 210–233. doi: [10.1016/0749-596X\(91\)90004-4](https://doi.org/10.1016/0749-596X(91)90004-4)
- Ferrer i Cancho, R., Hernández-Fernández, A., Lusseau, D., Agoramoorthy, G., Hsu, M. J., & Semple, S. (2013). Compression as a universal principle of animal behavior. *Cognitive science*, *37*(8), 1565–1578. doi: [10.1111/cogs.12061](https://doi.org/10.1111/cogs.12061)
- Fortson IV, B. W. (2004). *Indo-european language and culture: An introduction* (Vol. 30). Wiley-Blackwell.
- Fox Tree, J. E. (1999). Listening in on monologues and dialogues. *Discourse processes*, *27*(1), 35–53. doi: [10.1080/01638539909545049](https://doi.org/10.1080/01638539909545049)
- Fox Tree, J. E., & Mayer, S. A. (2008). Overhearing single and multiple perspectives. *Discourse Processes*, *45*(2), 160–179. doi: [10.1080/01638530701792867](https://doi.org/10.1080/01638530701792867)
- Fraser, C., Bellugi, U., & Brown, R. (1963). Control of grammar in imitation, comprehension, and production. *Journal of verbal learning and verbal behavior*, *2*(2), 121–135. doi: [10.1016/S0022-5371\(63\)80076-6](https://doi.org/10.1016/S0022-5371(63)80076-6)
- Fusaroli, R., & Tylén, K. (2012). Carving language for social coordination: a dynamical approach. *Interaction studies*, *13*(1), 103–124. doi: [10.1075/is.13.1.07fus](https://doi.org/10.1075/is.13.1.07fus)
- Fussell, S. R., & Krauss, R. M. (1989). The effects of intended audience on message production and comprehension: Reference in a common ground framework. *Journal of Experimental Social Psychology*, *25*(3), 203–219. doi: [10.1016/0022-1031\(89\)90019-X](https://doi.org/10.1016/0022-1031(89)90019-X)
- Futrell, R., Mahowald, K., & Gibson, E. (2015). Quantifying word order freedom in dependency corpora. In *Proceedings of the third international conference on dependency linguistics (depling)* (pp. 91–100). Retrieved from <https://www.aclweb.org/anthology/W15-2112>
- Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive science*, *29*(5), 737–767. doi: [10.1207/s15516709cog0000_34](https://doi.org/10.1207/s15516709cog0000_34)

- Galantucci, B., & Garrod, S. (2011). Experimental semiotics: a review. *Frontiers in human neuroscience*, 5, 11. doi: [10.3389/fnhum.2011.00011](https://doi.org/10.3389/fnhum.2011.00011)
- Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27(2), 181–218. doi: [10.1016/0010-0277\(87\)90018-7](https://doi.org/10.1016/0010-0277(87)90018-7)
- Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, 53(3), 181–215. doi: [10.1016/0010-0277\(94\)90048-5](https://doi.org/10.1016/0010-0277(94)90048-5)
- Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007). Foundations of representation: where might graphical symbol systems come from? *Cognitive Science*, 31(6), 961–987. doi: [10.1080/03640210701703659](https://doi.org/10.1080/03640210701703659)
- Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, 1(2), 292–304. doi: [10.1111/j.1756-8765.2009.01020.x](https://doi.org/10.1111/j.1756-8765.2009.01020.x)
- Gell-Mann, M. (2003). Effective complexity. In M. Gell-Mann & S. Lloyd (Eds.), *Nonextensive entropy – interdisciplinary applications* (pp. 387–398). Oxford: Oxford University Press.
- Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. technical report no. 257.
- Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. *Image, language, brain*, 95–126. Retrieved from https://tedlab.mit.edu/tedlab_website/researchpapers/Gibson_2000_DLT.pdf
- Gil, D. (2000). Syntactic categories, cross-linguistic variation and universal grammar. In *Approaches to the typology of word classes* (pp. 173–216). De Gruyter.
- Gil, D. (2009). How much grammar does it take to sail a boat? In G. Sampson, D. Gil, & P. Trudgill (Eds.), *Language complexity as an evolving variable*. Oxford University Press.
- Gilbert, R. (2007). The simultaneous manipulation of task complexity along planning time and [+/-here-and-now]: Effects on 12 oral production. In M. P. García Mayo (Ed.), *Investigating tasks in formal language learning* (Vol. 20, pp. 44–68). Bristol, UK: Multilingual Matters.
- Good, J. (2012). Typologizing grammatical complexities: or why creoles may be paradigm-

- matically simple but syntagmatically average. *Journal of Pidgin and Creole Languages*, 27(1), 1–47. doi: [10.1075/jpcl.27.1.01goo](https://doi.org/10.1075/jpcl.27.1.01goo)
- Good, J. (2015). Paradigmatic complexity in pidgins and creoles. *Word Structure*, 8(2), 184–227. doi: [10.3366/word.2015.0081](https://doi.org/10.3366/word.2015.0081)
- Greenberg, J. H. (1966). *Universals of language*. MIT press.
- Grice, H. P. (1957). Meaning. *The Philosophical review*, 66(3), 377–388. doi: [10.2307/2182440](https://doi.org/10.2307/2182440)
- Griffiths, T. L., Kalish, M. L., & Lewandowsky, S. (2008). Theoretical and empirical evidence for the impact of inductive biases on cultural evolution. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 363(1509), 3503–3514. doi: [10.1098/rstb.2008.0146](https://doi.org/10.1098/rstb.2008.0146)
- Grünwald, P. D., Myung, I. J., & Pitt, M. A. (2005). *Advances in minimum description length: Theory and applications*. Cambridge, MA: MIT press.
- Guiraud, P. (1954). *Les caractères statistiques du vocabulaire*. Presses universitaires de France.
- Haiman, J. (1983). Iconic and economic motivation. *Language*, 781–819. doi: [10.2307/413373](https://doi.org/10.2307/413373)
[Cite this publication](#)
- Håkansson, G. (1995). Syntax and morphology in language attrition: A study of five bilingual expatriate swedes. *International Journal of Applied Linguistics*, 5(2), 153–169.
- Hall, C. J. (1992). *Morphology and mind*. London/New York: Routledge.
- Hamming, R. W. (1950). Error detecting and error correcting codes. *Bell System technical journal*, 29(2), 147–160. doi: [10.1002/j.1538-7305.1950.tb00463.x](https://doi.org/10.1002/j.1538-7305.1950.tb00463.x)
- Hare, M., & Elman, J. L. (1995). Learning and morphological change. *Cognition*, 56(1), 61–98. doi: [10.1016/0010-0277\(94\)00655-5](https://doi.org/10.1016/0010-0277(94)00655-5)
- Harrington, J. (2006). An acoustic analysis of happy-tensing in the queen's christmas broadcasts. *Journal of Phonetics*, 34(4), 439–457. doi: [10.1016/j.wocn.2005.08.001](https://doi.org/10.1016/j.wocn.2005.08.001)
- Haspelmath, M. (2015). Framework-free grammatical theory. In B. Heine & H. Narrog (Eds.), *The oxford handbook of linguistic analysis* (pp. 287–311). Oxford: Oxford University Press. doi: [10.1093/oxfordhb/9780199544004.013.0014](https://doi.org/10.1093/oxfordhb/9780199544004.013.0014)
- Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

- Hawkins, J. A. (2009). An efficiency theory of complexity and related phenomena. In G. Sampson, D. Gil, & P. Trudgill (Eds.), *Language complexity as an evolving variable* (Vol. 13). Oxford: Oxford University Press.
- Heine, B., & Kuteva, T. (2007). *The genesis of grammar: A reconstruction* (Vol. 9). Oxford: Oxford University Press.
- Henrich, J., & Boyd, R. (2002). On modeling cognition and culture: Why cultural evolution does not require replication of representations. *Journal of cognition and culture*, 2(2), 87–112. doi: [10.1163/156853702320281836](https://doi.org/10.1163/156853702320281836)
- Hockett, C. F. (1960a). *A course in modern linguistics*. Macmillan.
- Hockett, C. F. (1960b). The origin of speech. *Scientific American*, 203, 88–111. doi: [10.1038/scientificamerican0960-88](https://doi.org/10.1038/scientificamerican0960-88)
- Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: Knowing when and knowing how to adjust utterances to addressees. *Journal of Memory and Language*, 47(4), 589–606. doi: [10.1016/S0749-596X\(02\)00019-0](https://doi.org/10.1016/S0749-596X(02)00019-0)
- Hout, R., & Vermeer, A. (2007, 01). Comparing measures of lexical richness. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (p. 93-116). Cambridge: Cambridge University Press.
- Hudson Kam, C. L. (2015). The impact of conditioning variables on the acquisition of variation in adult and child learners. *Language*, 91(4), 906–937. doi: [10.1353/lan.2015.0051](https://doi.org/10.1353/lan.2015.0051)
- Hudson Kam, C. L., & Chang, A. (2009). Investigating the cause of language regularization in adults: Memory constraints or learning effects? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(3), 815–21. doi: [10.1037/a0015097](https://doi.org/10.1037/a0015097)
- Hudson Kam, C. L., & Newport, E. (2001). Regularization during creolization: The learning of inconsistent linguistic input. In *26th annual boston university conference on language development, boston, ma*.
- Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language learning and development*, 1(2), 151–195. doi: [10.1080/15475441.2005.9684215](https://doi.org/10.1080/15475441.2005.9684215)
- Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive psychology*, 59(1), 30–66. doi: [10.1016/j.cogpsych.2009.01.001](https://doi.org/10.1016/j.cogpsych.2009.01.001)

- Hulk, A., & Cornips, L. (2006). Neuter gender and interface vulnerability in child I2/211 dutch. In S. Unsworth, T. Parodi, A. Sorace, & M. Young-Scholten (Eds.), *Paths of development in I1 and I2 acquisition: In honor of bonnie d. schwartz* (pp. 107–134). Amsterdam: John Benjamins. doi: [10.1075/lald.39.06hul](https://doi.org/10.1075/lald.39.06hul)
- Hupet, M., & Chantraine, Y. (1992). Changes in repeated references: Collaboration or repetition effects? *Journal of psycholinguistic research*, 21(6), 485–496. doi: [10.1007/BF01067526](https://doi.org/10.1007/BF01067526)
- Jaccard, P. (1912). The distribution of the flora in the alpine zone. *New phytologist*, 11(2), 37–50. doi: [10.1111/j.1469-8137.1912.tb05611.x](https://doi.org/10.1111/j.1469-8137.1912.tb05611.x)
- Jansson, F., Parkvall, M., & Strimling, P. (2015). Modeling the evolution of creoles. *Language Dynamics and Change*, 5(1), 1–51. doi: [10.1163/22105832-00501005](https://doi.org/10.1163/22105832-00501005)
- Kanwal, J., Smith, K., Culbertson, J., & Kirby, S. (2017). Zipfs law of abbreviation and the principle of least effort: Language users optimise a miniature lexicon for efficient communication. *Cognition*, 165, 45–52. doi: [10.1016/j.cognition.2017.05.001](https://doi.org/10.1016/j.cognition.2017.05.001)
- Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, 336(6084), 1049–1054. doi: [10.1126/science.1218811](https://doi.org/10.1126/science.1218811)
- Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30(1/2), 81–93. doi: [10.2307/2332226](https://doi.org/10.2307/2332226)
- Kendall, M. G. (1945). The treatment of ties in ranking problems. *Biometrika*, 33(3), 239–251. doi: [10.1093/biomet/33.3.239](https://doi.org/10.1093/biomet/33.3.239)
- Kerswill, P. (1996). Children, adolescents, and language change. *Language variation and change*, 8(2), 177–202. doi: [10.1017/S0954394500001137](https://doi.org/10.1017/S0954394500001137)
- Kerswill, P., & Williams, A. (2000). Creating a new town koine: children and language change in milton keynes. *Language in society*, 29(1), 65–115. doi: [10.1017/S0047404500001020](https://doi.org/10.1017/S0047404500001020)
- Kirby, S. (1997). Competing motivations and emergence: explaining implicational hierarchies. *Language Typology*, 1(1), 5–32. doi: [10.1515/lity.1997.1.1.5](https://doi.org/10.1515/lity.1997.1.1.5)
- Kirby, S. (1999). *Function, selection, and innateness: The emergence of language universals*. Oxford: OUP Oxford.
- Kirby, S. (2002). Learning, bottlenecks and the evolution of recursive syntax. In T. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational*

- models* (pp. 173–204). Cambridge, UK: Cambridge University Press.
- Kirby, S., Cornish, H., & Smith, K. (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. doi: [10.1073/pnas.0707835105](https://doi.org/10.1073/pnas.0707835105)
- Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current opinion in neurobiology*, *28*, 108–114. doi: [10.1016/j.conb.2014.07.014](https://doi.org/10.1016/j.conb.2014.07.014)
- Kirby, S., & Hurford, J. R. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In *Simulating the evolution of language* (pp. 121–147). Springer. doi: [10.1007/978-1-4471-0663-0_6](https://doi.org/10.1007/978-1-4471-0663-0_6)
- Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102. doi: [10.1016/j.cognition.2015.03.016](https://doi.org/10.1016/j.cognition.2015.03.016)
- Klein, W., & Perdue, C. (1997). The basic variety (or: Couldn't natural languages be much simpler?). *Second language research*, *13*(4), 301–347. doi: [10.1191/026765897666879396](https://doi.org/10.1191/026765897666879396)
- Kouwenberg, S., & Singler, J. V. (2009). *The handbook of pidgin and creole studies*. Oxford: Wiley-Blackwell.
- Krauss, R. M., Vivekananthan, P., & Weinheimer, S. (1968). “inner speech” and “external speech”: characteristics and communication effectiveness of socially and nonsocially encoded messages. *Journal of Personality and Social Psychology*, *9*(4), 295–300. doi: [10.1037/h0026091](https://doi.org/10.1037/h0026091)
- Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, *1*(1-12), 113–114. doi: [10.3758/BF03342817](https://doi.org/10.3758/BF03342817)
- Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of personality and social psychology*, *4*(3), 343–46. doi: [10.1037/h0023705](https://doi.org/10.1037/h0023705)
- Kusters, W. (2003). *Linguistic complexity*. Netherlands Graduate School of Linguistics.
- Kusters, W. (2008). Complexity in linguistic theory, language learning and language change. In M. Miestamo, S. Kaius, & F. Karlsson (Eds.), *Language complexity: Typology, contact, change* (pp. 3–22). Amsterdam: John Benjamins.
- Kuteva, T. (2008). On the frills of grammaticalization. In M. J. López & E. Seoane (Eds.),

- Rethinking grammaticalization: New perspectives* (Vol. 76, pp. 189–217). Amsterdam: John Benjamins. doi: [10.1075/tsl.76.10kut](https://doi.org/10.1075/tsl.76.10kut)
- Kuznetsova, A., Brockhoff, P., & Christensen, R. (2014). *Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). r package version 2.0-3*. Retrieved from <https://cran.r-project.org/web/packages/lmerTest/index.html>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). *lmerTest: Tests in linear mixed effects models [Computer software manual]*. Retrieved from <http://CRAN.R-project.org/package=lmerTest> (R package version 2.0-29)
- Labov, W. (1972). *Sociolinguistic patterns* (No. 4). University of Pennsylvania Press.
- Lachlan, R. F., & Nowicki, S. (2015). Context-dependent categorical perception in a songbird. *Proceedings of the National Academy of Sciences*, *112*(6), 1892–1897. doi: [10.1073/pnas.1410844112](https://doi.org/10.1073/pnas.1410844112)
- Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive development*, *3*(3), 299–321. doi: [10.1016/0885-2014\(88\)90014-7](https://doi.org/10.1016/0885-2014(88)90014-7)
- Lefebvre, C., White, L., & Jourdan, C. (2006). *L2 acquisition and creole genesis: dialogues* (Vol. 42). Amsterdam: John Benjamins.
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady* (Vol. 10, pp. 707–710). Retrieved from <http://adsabs.harvard.edu/abs/1966SPhD...10..707L> (Provided by the SAO/NASA Astrophysics Data System)
- Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.
- Lewis, D. (1968). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.
- Lewis, D. (1969). *Convention: A philosophical study*. Basil Blackwell/Harvard University Press.
- Li, M., & Vitányi, P. (1997). *An introduction to kolmogorov complexity and its applications*. Heidelberg: Springer.
- Lightfoot, D. (2010). Language acquisition and language change. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(5), 677–684. doi: [10.1002/wcs.39](https://doi.org/10.1002/wcs.39)

- Lloyd, S. (2001). Measures of complexity: a nonexhaustive list. *IEEE Control Systems Magazine*, 21(4), 7–8. doi: [10.1109/MCS.2001.939938](https://doi.org/10.1109/MCS.2001.939938)
- Long, M. (1985). Input and second language acquisition theory. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 377–393). Rowley, MA: Newbury House.
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4), 474–496. doi: [10.1075/ijcl.15.4.02lu](https://doi.org/10.1075/ijcl.15.4.02lu)
- Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PloS one*, 5(1), e8559. doi: [10.1371/journal.pone.0008559](https://doi.org/10.1371/journal.pone.0008559)
- Lupyan, G., & Dale, R. (2015). The role of adaptation in understanding linguistic diversity. In R. D. Busser & R. J. LaPolla (Eds.), *Language structure and environment: Social, cultural, and natural factors.cognitive linguistic studies in cultural contexts*, 6. John Benjamins. doi: [10.1075/clsc.6.11lup](https://doi.org/10.1075/clsc.6.11lup)
- Marantz, A. (2005). Generative linguistics within the cognitive neuroscience of language. *The Linguistic Review*, 22(2-4), 429–445. doi: [10.1515/tlir.2005.22.2-4.429](https://doi.org/10.1515/tlir.2005.22.2-4.429)
- Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F., & Clahsen, H. (1992). Overregularization in language acquisition. *Monographs of the society for research in child development*, 57(4), i–178. doi: [10.2307/1166115](https://doi.org/10.2307/1166115)
- Markus, J., Mundy, P., Morales, M., Delgado, C. E., & Yale, M. (2000). Individual differences in infant skills as predictors of child-caregiver joint attention and language. *Social Development*, 9(3), 302–315. doi: [10.1111/1467-9507.00127](https://doi.org/10.1111/1467-9507.00127)
- Martinet, A. (1967). *Eléments de linguistique générale*. Paris: Armand Colin.
- Massip-Bonet, À., & Bastardas-Boada, A. (2012). *Complexity perspectives on language, communication and society*. Heidelberg: Springer.
- Mcallister, J., Potts, A., Mason, K., & Marchant, G. (1994). Word duration in monologue and dialogue speech. *Language and Speech*, 37(4), 393–405. doi: [10.1177/002383099403700404](https://doi.org/10.1177/002383099403700404)
- McCauley, S. M., & Christiansen, M. H. (2011). Learning simple statistics for language comprehension and production: The cappuccino model. In *Proceedings of the cognitive science society* (Vol. 33). Austin, TX. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.208.1404&rep=rep1&type=pdf>

- McWhorter, J. (2001). The worlds simplest grammars are creole grammars. *Linguistic typology*, 5(2/3), 125–166. doi: [10.1515/lity.2001.001](https://doi.org/10.1515/lity.2001.001)
- McWhorter, J. (2005). *Defining creole*. Oxford: Oxford University Press.
- McWhorter, J. (2007). *Language interrupted: Signs of non-native acquisition in standard language grammars*. Oxford: Oxford University Press. doi: [10.1093/acprof:oso/9780195309805.001.0001](https://doi.org/10.1093/acprof:oso/9780195309805.001.0001)
- Mesoudi, A. (2011). Variable cultural acquisition costs constrain cumulative cultural evolution. *PloS one*, 6(3), e18239. doi: [10.1371/journal.pone.0018239](https://doi.org/10.1371/journal.pone.0018239)
- Mesoudi, A., & Whiten, A. (2008). The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 363(1509), 3489–3501. doi: [10.1098/rstb.2008.0129](https://doi.org/10.1098/rstb.2008.0129)
- Michel, M. C. (2011). Effects of task complexity and interaction on L2 performance. *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance*, 2, 141–173. doi: [10.1075/tblt.2.12ch6](https://doi.org/10.1075/tblt.2.12ch6)
- Miestamo, M. (2006a). On the complexity of standard negation. In M. Suominen et al. (Eds.), *A man of measure: Festschrift in honour of fred karlsson on his 60th birthday* (pp. 345–356). Turku: The Linguistic Association of Finland.
- Miestamo, M. (2006b). On the feasibility of complexity metrics. In *Finest linguistics, proceedings of the annual finnish and estonian conference of linguistics, tallinn, may 6–7, 2004* (pp. 11–26). Retrieved from https://www.academia.edu/21761885/On_the_Feasibility_of_Complexity_Metrics
- Miestamo, M. (2008). Grammatical complexity in a cross-linguistic perspective. In M. Miestamo, S. Kaius, & F. Karlsson (Eds.), *Language complexity: Typology, contact, change* (pp. 23–41). Amsterdam: John Benjamins.
- Miestamo, M. (2009). Implicational hierarchies and grammatical complexity. In G. Sampson, D. Gil, & P. Trudgill (Eds.), *Language complexity as an evolving variable* (Vol. 13). Oxford: Oxford University Press.
- Mikulecky, D. C. (2001). Network thermodynamics and complexity: a transition to relational systems theory. *Computers & chemistry*, 25(4), 369–391. doi: [10.1016/S0097-8485\(01\)00072-9](https://doi.org/10.1016/S0097-8485(01)00072-9)
- Miller, G. A. (1958). Free recall of redundant strings of letters. *Journal of Experimental*

- Psychology*, 56(6), 485. doi: [10.1037/h0044933](https://doi.org/10.1037/h0044933)
- Miller, K. L., & Schmitt, C. (2012). Variable input and the acquisition of plural morphology. *Language Acquisition*, 19(3), 223–261. doi: [10.1080/10489223.2012.685026](https://doi.org/10.1080/10489223.2012.685026)
- Millikan, R. G. (1998). Language conventions made simple. *The Journal of Philosophy*, 95(4), 161–180. doi: [10.2307/2564683](https://doi.org/10.2307/2564683)
- Montague, R. (1970). Universal grammar. *Theoria*, 36(3), 373–398. doi: [10.1111/j.1755-2567.1970.tb00434.x](https://doi.org/10.1111/j.1755-2567.1970.tb00434.x)
- Montemurro, M. A., & Zanette, D. H. (2011). Universal entropy of word ordering across linguistic families. *PLoS One*, 6(5), e19875.
- Montrul, S. A. (2004). *The acquisition of spanish: Morphosyntactic development in monolingual and bilingual l1 acquisition and adult l2 acquisition* (Vol. 37). John Benjamins.
- Motamedi, Y., Schouwstra, M., Smith, K., Culbertson, J., & Kirby, S. (under revision). Evolving artificial sign languages in the lab: from improvised gesture to systematic sign.
- Mufwene, S. S. (2008). *Language evolution: Contact, competition and change*. London: Bloomsbury.
- Mufwene, S. S. (2012). The emergence of complexity in language: An evolutionary perspective. In À. Massip-Bonet & A. Bastardas-Boada (Eds.), *Complexity perspectives on language, communication and society* (pp. 197–218). Heidelberg: Springer.
- Mufwene, S. S. (2013). Language as technology some questions that evolutionary linguistics should address. *In Search of Universal Grammar: From Old Norse to Zoque*, 202, 327. doi: [0.1075/la.202.22muf](https://doi.org/0.1075/la.202.22muf)
- Mufwene, S. S., Coupé, C., & Pellegrino, F. (2017). *Complexity in language: developmental and evolutionary perspectives*. Cambridge: Cambridge University Press. doi: [10.1017/9781107294264](https://doi.org/10.1017/9781107294264)
- Mühlhäusler, P. (1977). *Pidginisation and simplification of language*. Canberra: Pacific Linguistics.
- Murfitt, T., & McAllister, J. (2001). The effect of production variables in monolog and dialog on comprehension by novel listeners. *Language and speech*, 44(3), 325–350. doi: [10.1177/00238309010440030201](https://doi.org/10.1177/00238309010440030201)
- Navarro, D. J., Perfors, A., Kary, A., Brown, S., & Donkin, C. (2017). When extremists win: on the behavior of iterated learning chains when priors are heteroge-

- neous. In *Proceedings of the 38th annual conference of the cognitive science society* (pp. 847–852). Retrieved from https://compcogscisydney.org/publications/NavarroPKBD_sheep.pdf
- Newmeyer, F. J., & Preston, L. B. (2014). *Measuring grammatical complexity*. Oxford: Oxford University Press.
- Newport, E. L. (1999). Reduced input in the acquisition of signed languages: Contributions to the study of creolization. In M. DeGraff (Ed.), *Language creation and language change: Creolization, diachrony, and development*. Cambridge, MA: The MIT Press.
- Nichols, J. (1992). *Linguistic diversity in space and time*. Chicago, IL: University of Chicago Press.
- Nichols, J. D. (1995). The ojibwe verb in broken oghibbeway. *Amsterdam Creole Studies*, 12, 1–18.
- Niyogi, P., & Berwick, R. C. (2009). The proper treatment of language acquisition and change in a population setting. *Proceedings of the National Academy of Sciences*, 106(25), 10124–10129. doi: [10.1073 / pnas.0903993106](https://doi.org/10.1073/pnas.0903993106)
- Nowak, I., & Baggio, G. (2016). The emergence of word order and morphology in compositional languages via multigenerational signaling games. *Journal of Language Evolution*, 137–150. doi: [10.1093/jole/lzw007](https://doi.org/10.1093/jole/lzw007)
- Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.
- O'Donnell, T. J. (2015). *Productivity and reuse in language: A theory of linguistic computation and storage*. Cambridge, MA.
- Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences*, 106(51), 22026–22031. doi: [10.1073/pnas.0908118106](https://doi.org/10.1073/pnas.0908118106)
- Page, S. E. (2010). *Diversity and complexity*. Princeton University Press.
- Pagin, P. (2012). Communication and the complexity of semantics. In *The oxford handbook of compositionality*. Oxford University Press.
- Pagin, P. (2013). Compositionality, complexity, and evolution. Retrieved from http://www.ling.su.se/polopoly_fs/1.142249.1376032271!/menu/standard/file/Pagin_KVA_Symposium_20130808.pdf

- Pagin, P., & Westerståhl, D. (2010). Compositionality i: Definitions and variants. *Philosophy Compass*, 5(3), 250–264. doi: [10.1111/j.1747-9991.2009.00228.x](https://doi.org/10.1111/j.1747-9991.2009.00228.x)
- Parkvall, M. (2008). The simplicity of creoles in a cross-linguistic perspective. In (pp. 265–285). doi: [10.1075/slcs.94.17par](https://doi.org/10.1075/slcs.94.17par)
- Peirce, J. W. (2007). Psychopy–psychophysics software in python. *Journal of neuroscience methods*, 162(1), 8–13. doi: [10.1016/j.jneumeth.2006.11.017](https://doi.org/10.1016/j.jneumeth.2006.11.017)
- Peirce, J. W. (2009). Generating stimuli for neuroscience using psychopy. *Frontiers in Neuroinformatics*, 2(10). doi: [10.3389/neuro.11.010.2008](https://doi.org/10.3389/neuro.11.010.2008)
- Perfors, A. (2012a). Probability matching vs over-regularization in language: Participant behavior depends on their interpretation of the task.. Retrieved from <https://mindmodeling.org/cogsci2012/papers/0155/paper0155.pdf>
- Perfors, A. (2012b). When do memory limitations lead to regularization? an experimental and computational investigation. *Journal of Memory and Language*, 67(4), 486–506. doi: [10.1016/j.jml.2012.07.009](https://doi.org/10.1016/j.jml.2012.07.009)
- Perfors, A. (2016). Adult regularization of inconsistent input depends on pragmatic factors. *Language Learning and Development*, 12(2), 138–155.
- Perfors, A., & Burns, N. (2010). Adult language learners under cognitive load do not over-regularize like children.. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.707.1474&rep=rep1&type=pdf>
- Perfors, A., & Navarro, D. J. (2014). Language evolution can be shaped by the structure of the world. *Cognitive science*, 38(4), 775–793. doi: [10.1111/cogs.12102](https://doi.org/10.1111/cogs.12102)
- Perruchet, P., & Desauty, S. (2008). A role for backward transitional probabilities in word segmentation? *Memory & Cognition*, 36(7), 1299–1305. doi: [10.3758/MC.36.7.1299](https://doi.org/10.3758/MC.36.7.1299)
- Petrov, S., Barrett, L., Thibaux, R., & Klein, D. (2006). Learning accurate, compact, and interpretable tree annotation. In *Proceedings of the 21st international conference on computational linguistics and the 44th annual meeting of the association for computational linguistics* (pp. 433–440). doi: [10.3115/1220175.1220230](https://doi.org/10.3115/1220175.1220230)
- Petrov, S., & Klein, D. (2007, April). Improved inference for unlexicalized parsing. In *Human language technologies 2007: The conference of the north american chapter of the association for computational linguistics; proceedings of the main conference* (pp. 404–411). Rochester, New York: Association for Computational Linguistics. Retrieved from

- <http://www.aclweb.org/anthology/N07-1051>
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, 27(02), 169–190.
- Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation*, 4(2-3), 203–228. doi: [10.1007/s11168-006-9004-0](https://doi.org/10.1007/s11168-006-9004-0)
- Pullum, G. K., & Scholz, B. C. (2007). Systematicity and natural language syntax. *Croatian Journal of Philosophy*, 7(21), 375–402. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.97.7486>
- Pygame Community. (2009). *Pygame - python game development*. <http://pygame.org/>.
- R Core Team. (2000). *R language definition*. Available from CRAN sites. Retrieved from cran.r-project.org
- R Core Team. (2015). *R: A language and environment for statistical computing [Computer software manual]*. Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Raymond, W. D., Healy, A. F., McDonnell, S., & Healy, C. A. (2009). Acquisition of morphological variation: The case of the english definite article. *Language and Cognitive Processes*, 24(1), 89–119. doi: [10.1080/01690960802075489](https://doi.org/10.1080/01690960802075489)
- Real, F., & Griffiths, T. L. (2009). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, 111(3), 317–328. doi: [10.1016/j.cognition.2009.02.012](https://doi.org/10.1016/j.cognition.2009.02.012)
- Regier, T., Kemp, C., & Kay, P. (2015). Word meanings across languages support efficient communication. *The handbook of language emergence*, 87, 237. doi: [10.1111/b.9781118301753.2015.00013.x](https://doi.org/10.1111/b.9781118301753.2015.00013.x)
- Rescher, N. (1998). *Complexity: A philosophical overview*. Transaction Publishers.
- Richards, J. C. (1974). *Error analysis: Perspectives on second language acquisition*. London: Longman.
- Richards, J. C. (1975). Simplification: A strategy in the adult acquisition of a foreign language: An example from Indonesian/Malay. *Language Learning*, 25(1), 115–126. doi: [10.1111/j.1467-1770.1975.tb00112.x](https://doi.org/10.1111/j.1467-1770.1975.tb00112.x)
- Roché, M. (1992). Le masculin est-il plus productif que le féminin? *Langue française*(96), 113–124.

- Ross, D. S., & Newport, E. L. (1996). The development of language from non-native linguistic input. In *Proceedings of the 20th annual boston university conference on language development* (Vol. 2, pp. 634–645). Retrieved from http://wearables.cc.gatech.edu/paper_of_week/Ross_and_Newport_1996.pdf
- Saldana, C. (2017). Data from "Simplifying linguistic complexity: Culture and cognition in language evolution", PhD dissertation, 2014-2017 [dataset]. The University of Edinburgh. Centre for Language Evolution. doi: [10.7488/ds/2235](https://doi.org/10.7488/ds/2235)
- Sankoff, G., & Laberge, S. (1974). On the acquisition of native speakers by a language. *Pidgins and creoles: Current trends and prospects*, 73–84. doi: [10.9783/9781512809589-014](https://doi.org/10.9783/9781512809589-014)
- Sasse, H.-J. (2001). Typological changes in language obsolescence. In M. Haspelmath, E. König, W. Oesterreicher, & W. Raible (Eds.), *Language typology and language universals: An international handbook* (pp. 1668–1677). Berlin: De Gruyter.
- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics bulletin*, 2(6), 110–114. doi: [10.2307/3002019](https://doi.org/10.2307/3002019)
- Schelling, T. C. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.
- Scherer, N. J., & Olswang, L. B. (1984). Role of mothers expansions in stimulating childrens language production. *Journal of Speech and Hearing Research*, 27(3), 387–396. doi: [10.1044/jshr.2703.387](https://doi.org/10.1044/jshr.2703.387)
- Schilling-Estes, N., & Wolfram, W. (1994). Convergent explanation and alternative regularization patterns: Were/weren't leveling in a vernacular english variety. *Language variation and change*, 6(03), 273–302. doi: [10.1017/S0954394500001691](https://doi.org/10.1017/S0954394500001691)
- Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive psychology*, 21(2), 211–232. doi: [10.1016/0010-0285\(89\)90008-X](https://doi.org/10.1016/0010-0285(89)90008-X)
- Schoneberger, T. (2010). Three myths from the language acquisition literature. *The Analysis of verbal behavior*, 26(1), 107–131. doi: [10.1007/BF03393086](https://doi.org/10.1007/BF03393086)
- Schumacher, R. A., Pierrehumbert, J. B., & LaShell, P. (2014). Reconciling inconsistency in encoded morphological distinctions in an artificial language. In *Proceedings of the 36th annual conference of the cognitive science society*.
- Scott-Phillips, T. C., & Kirby, S. (2010). Language evolution in the laboratory. *Trends in cognitive sciences*, 14(9), 411–417. doi: [10.1016/j.tics.2010.06.006](https://doi.org/10.1016/j.tics.2010.06.006)
- Selten, R., & Warglien, M. (2007). The emergence of simple languages in an experimental

- coordination game. *Proceedings of the National Academy of Sciences*, 104(18), 7361–7366. doi: [10.1073/pnas.0702077104](https://doi.org/10.1073/pnas.0702077104)
- Senghas, A., & Coppola, M. (2001). Children creating language: How nicaraguan sign language acquired a spatial grammar. *Psychological science*, 12(4), 323–328. doi: [10.1111/1467-9280.00359](https://doi.org/10.1111/1467-9280.00359)
- Senghas, A., Kita, S., & Özyürek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in nicaragua. *Science*, 305(5691), 1779–1782. doi: [10.1126/science.1100199](https://doi.org/10.1126/science.1100199)
- Senghas, A., Newport, E. L., & Supalla, T. (1997). Argument structure in nicaraguan sign language: The emergence of grammatical devices. In *Proceedings of the 21st annual boston university conference on language development*. Retrieved from https://www.researchgate.net/publication/237295284_Argument_Structure_in_Nicaraguan_Sign_Language_The_Emergence_of_Grammatical_Devices
- Shuy, R. W., Wolfram, W., & Riley, W. K. (1967). *Linguistic correlates of social stratification in detroit speech*. Technical Report U. S. Office of Education Final Report No. 6-1347.
- Siegel, J. (1997). Mixing, leveling, and pidgin/creole development. In A. K. Spears & D. Winford (Eds.), *The structure and status of pidgins and creoles* (pp. 111–149). Amsterdam/Philadelphia: John Benjamins. doi: [10.1075/cll.19.09sie](https://doi.org/10.1075/cll.19.09sie)
- Siegel, J. (2004). Morphological simplicity in pidgins and creoles. *Journal of Pidgin and Creole languages*, 19(1), 139–162. doi: [10.1075/jpocl.19.1.06sie](https://doi.org/10.1075/jpocl.19.1.06sie)
- Siegel, J. (2006). Links between sla and creole studies: Past and present. *Language Acquisition and Language Disorders*, 42, 15.
- Silvey, C., Kirby, S., & Smith, K. (2015). Word meanings evolve to selectively preserve distinctions on salient dimensions. *Cognitive science*, 39(1), 212–226. doi: [10.1111/cogs.12150](https://doi.org/10.1111/cogs.12150)
- Simon, H. A. (1996). *The sciences of the artificial*. Cambridge, MA: MIT press.
- Singleton, J. L., & Newport, E. L. (2004). When learners surpass their models: The acquisition of american sign language from inconsistent input. *Cognitive psychology*, 49(4), 370–407. doi: [10.1016/j.cogpsych.2004.05.001](https://doi.org/10.1016/j.cogpsych.2004.05.001)
- Sinnemäki, K., et al. (2011). *Language universals and linguistic complexity: Three case studies in core argument marking* (Unpublished doctoral dissertation). University of Helsinki.

- Slobin, D. I. (1966). The acquisition of russian as a native language. *The genesis of language: A psycholinguistic approach*, 129–148.
- Slobin, D. I. (1986). The acquisition and use of relative clauses in turkic and indo-european languages. *Studies in Turkish linguistics*, 8, 273. doi: [10.1075/tsl.8.16slo](https://doi.org/10.1075/tsl.8.16slo)
- Smith, C. B., Adamson, L. B., & Bakeman, R. (1988). Interactional predictors of early language. *First Language*, 8(23), 143–156. doi: [10.1177/014272378800802304](https://doi.org/10.1177/014272378800802304)
- Smith, K. (2004). The evolution of vocabulary. *Journal of theoretical biology*, 228(1), 127–142. doi: [10.1016/j.jtbi.2003.12.016](https://doi.org/10.1016/j.jtbi.2003.12.016)
- Smith, K. (2009). Iterated learning in populations of bayesian agents. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 31). Retrieved from <https://pdfs.semanticscholar.org/a4ea/2ffa35a032830502f8606db091d572db1aeb.pdf>
- Smith, K., Brighton, H., & Kirby, S. (2003). Complex systems in language evolution: the cultural emergence of compositional structure. *Advances in Complex Systems*, 6(04), 537–558. doi: [10.1142/S0219525903001055](https://doi.org/10.1142/S0219525903001055)
- Smith, K., & Kirby, S. (2008). Cultural evolution: implications for understanding the human language faculty and its evolution. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 363(1509), 3591–3603. doi: [10.1098/rstb.2008.0145](https://doi.org/10.1098/rstb.2008.0145)
- Smith, K., & Kirby, S. (2012). Compositionality and linguistic evolution. In *The oxford handbook of compositionality*. Oxford University Press.
- Smith, K., Perfors, A., Fehér, O., Samara, A., Swoboda, K., & Wonnacott, E. (2017). Language learning, language use and the evolution of linguistic variation. *Phil. Trans. R. Soc. B*, 372(1711), 20160051. doi: [10.1098/rstb.2016.0051](https://doi.org/10.1098/rstb.2016.0051)
- Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, 116(3), 444–449. doi: [10.1016/j.cognition.2010.06.004](https://doi.org/10.1016/j.cognition.2010.06.004)
- Spears, A. K. (2008). Pidgins/creoles and african american english. In S. Kouwenberg & J. V. Singler (Eds.), *The handbook of pidgin and creole studies* (pp. 512–42). Oxford: Blackwell. doi: [10.1002/9781444305982.ch21](https://doi.org/10.1002/9781444305982.ch21)
- Spike, M., Stadler, K., Kirby, S., & Smith, K. (2017). Minimal requirements for the emergence of learned signaling. *Cognitive science*, 41(3), 623–658. doi: [10.1111/cogs.12351](https://doi.org/10.1111/cogs.12351)
- Steels, L. (1995). A self-organizing spatial vocabulary. *Artificial life*, 2(3), 319–332. doi: [10.1162/artl.1995.2.319](https://doi.org/10.1162/artl.1995.2.319)

- Steels, L. (2012). Self-organization and selection in cultural language evolution. In L. Steels (Ed.), *Experiments in cultural language evolution* (p. 1-37). Amsterdam: John Benjamins. doi: [10.1075/ais.3.02ste](https://doi.org/10.1075/ais.3.02ste)
- Stefnsson, V. (1909). The eskimo trade jargon of herschel island [Bibliographical record]. *American Anthropologist*, *11*, 217-232. (This is the first publication dedicated to an Eskimo Pidgin and it is the only linguistic publication dedicated to an Eskimo Pidgin that is based on primary fieldwork. All examples are first-hand. The pidgin is representative of Eskimo Pidgins in the Western American Arctic in general.)
- Stevin, S. (1586). *De beghinselen der weeghconst*. Leiden.
- Strapp, C. M., Bleakney, D. M., Helmick, A. L., & Tonkovich, H. M. (2008). Developmental differences in the effects of negative and positive evidence. *First Language*, *28*(1), 35–53. doi: [10.1177/0142723707084840](https://doi.org/10.1177/0142723707084840)
- Suzuki, T. N., Wheatcroft, D., & Griesser, M. (2016). Experimental evidence for compositional syntax in bird calls. *Nature communications*, *7*. doi: [10.1038/ncomms10986](https://doi.org/10.1038/ncomms10986)
- Szabó, Z. G. (2012). The case for compositionality. In *The oxford handbook of compositionality*. Oxford University Press.
- Szmrecsany, B., & Kortmann, B. (2009). Between simplification and complexification: non-standard varieties of english around the world. In G. Sampson, D. Gil, & P. Trudgill (Eds.), *Language complexity as an evolving variable* (Vol. 13, pp. 64–79). Oxford: Oxford University Press.
- Theisen, C. A., Oberlander, J., & Kirby, S. (2010). Systematicity and arbitrariness in novel communication systems. *Interaction Studies*, *11*(1), 14–32. doi: [10.1075/is.11.1.08the](https://doi.org/10.1075/is.11.1.08the)
- Theisen-White, C., Kirby, S., & Oberlander, J. (2011). Integrating the horizontal and vertical cultural transmission of novel communication systems. In *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 956–961). Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?rep=rep1&type=pdf&doi=10.1.1.208.1489>
- Thomason, S. G. (1980). On interpreting the indian interpreter. *Language in Society*, *9*(2), 167–193.
- Tinits, P., Nölle, J., & Hartmann, S. (2017). Usage context influences the evolution of overspecification in iterated learning. *Journal of Language Evolution*. doi: [10.1093/jole/lzx011](https://doi.org/10.1093/jole/lzx011)

- Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child development*, 57(6), 1454–63. doi: [10.2307/1130423](https://doi.org/10.2307/1130423)
- Touchie, H. Y. (1986). Second language learning errors: Their types, causes, and treatment. *JALT journal*, 8(1), 75–80. Retrieved from <http://jalt-publications.org/jj/articles/1571-second-language-learning-errors-their-types-causes-and-treatment>
- Traxler, M. J., & Gernsbacher, M. A. (1992). Improving written communication through minimal feedback. *Language and Cognitive Processes*, 7(1), 1–22. doi: [10.1080/01690969208409378](https://doi.org/10.1080/01690969208409378)
- Trudgill, P. (1974). *The social differentiation of english in norwich* (Vol. 13). CUP Archive.
- Trudgill, P. (2011). *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: Oxford University Press.
- Vainikka, A., & Young-Scholten, M. (1996). The early stages in adult L2 syntax: additional evidence from romance speakers. *Second Language Research*, 12(2), 140–176. doi: [10.1177/026765839601200202](https://doi.org/10.1177/026765839601200202)
- van der Voort, H. (1997). New light on eskimo pidgins. In A. K. Spears & D. Winford (Eds.), *The structure and status of pidgins and creoles: Including selected papers from meetings of the society for pidgin and creole linguistics* (Vol. 19). Amsterdam: John Benjamins.
- van der Voort, H. (2013). Eskimo pidgin structure dataset. In S. M. Michaelis, P. Maurer, M. Haspelmath, & M. Huber (Eds.), *Atlas of pidgin and creole language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://apics-online.info/contributions/76>
- Vanhove, J. (2014, 20th August). *Calibrating p-values in 'flexible' piecewise regression models*. Retrieved from <https://janhove.github.io/analysis/2014/08/20/adjusted-pvalues-breakpoint-regression>.
- van Trijp, R. (2012). The evolution of case systems for marking event structure. In L. Steels (Ed.), *Experiments in cultural language evolution* (Vol. 3, pp. 169–205). Amsterdam: John Benjamins.
- van Trijp, R. (2013). Linguistic assessment criteria for explaining language change: A case study on syncretism in german definite articles. *Language Dynamics and Change*, 3(1), 105–132. doi: [10.1163/22105832-13030106](https://doi.org/10.1163/22105832-13030106)

- van Trijp, R. (2016). *The evolution of case grammar*. Berlin: Language Science Press.
- Velupillai, V. (2015). *Pidgins, creoles and mixed languages: An introduction* (Vol. 48). Amsterdam/Philadelphia: John Benjamins.
- Verhoef, T. (2012). The origins of duality of patterning in artificial whistled languages. *Language and Cognition*, 4(4), 357–380. doi: [10.1515/langcog-2012-0019](https://doi.org/10.1515/langcog-2012-0019)
- Weber, K., Christiansen, M. H., Petersson, K. M., Indefrey, P., & Hagoort, P. (2016). fmri syntactic and lexical repetition effects reveal the initial stages of learning a new language. *Journal of Neuroscience*, 36(26), 6872–6880. doi: [10.1523/JNEUROSCI.3180-15.2016](https://doi.org/10.1523/JNEUROSCI.3180-15.2016)
- Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of memory and language*, 31(2), 183–194. doi: [10.1016/0749-596X\(92\)90010-U](https://doi.org/10.1016/0749-596X(92)90010-U)
- Winford, D. (2003). *An introduction to contact linguistics*. Oxford: Wiley-Blackwell.
- Winters, J. (2017). *Context, cognition and communication in language* (Doctoral dissertation, the University of Edinburgh). Retrieved from <https://www.era.lib.ed.ac.uk/handle/1842/23546>
- Winters, J., Kirby, S., & Smith, K. (2015). Languages adapt to their contextual niche. *Language and Cognition*, 7(03), 415–449. doi: [10.1017/langcog.2014.35](https://doi.org/10.1017/langcog.2014.35)
- Winters, J., Kirby, S., & Smith, K. (2018). Contextual predictability shapes signal autonomy. *Cognition*, 176, 15–30.
- Wonnacott, E., & Newport, E. L. (2005). Novelty and regularization: The effect of novel instances on rule formation. In *BuclD 29: Proceedings of the 29th annual boston university conference on language development* (Vol. 29, pp. 663–673). Somerville: Cascadilla Press. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.222.7380&rep=rep1&type=pdf>
- Wray, A., & Grace, G. W. (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, 117(3), 543–578. doi: [10.1016/j.lingua.2005.05.005](https://doi.org/10.1016/j.lingua.2005.05.005)
- Yang, C. (2016). *The price of linguistic productivity: How children learn to break the rules of language*. Cambridge, MA: The MIT Press.
- Yip, M. J. (2006). The search for phonology in other species. *Trends in cognitive sciences*, 10(10), 442–446. doi: [10.1016/j.tics.2006.08.001](https://doi.org/10.1016/j.tics.2006.08.001)
- Yule, G. (1996). *The study of language*. 2nd. Cambridge: Cambridge University Press.

- Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Boston, MA: Addison-Wesley Press.
- Zuidema, W. H. (2003). How the poverty of the stimulus solves the poverty of the stimulus. In *Advances in neural information processing systems* (pp. 51–58). Retrieved from http://machinelearning.wustl.edu/mlpapers/paper_files/CS07.pdf

Appendix A

Relevant papers

A.1 Author contributions

This appendix contains a conference paper which was co-authored with Kenny Smith, Simon Kirby and Jennifer Culbertson and published on-line in the Conference Proceedings of the 39th Meeting of Cognitive Science Society. The paper contains Experiments 5 and 6 of Chapter 4 (sections 4.3 and 4.4 respectively). The experiments were conceived during supervision meetings with all co-authors present; I conducted the experiments and analyses; and all co-authors jointly contributed to the writing of the paper.

A.2 Saldana, Smith, Kirby & Culbertson (2017) Is the strength of regularisation uniform across linguistic levels? *Proceedings of the 39th Annual Conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Is the strength of regularisation behaviour uniform across linguistic levels?

Carmen Saldana (C.C.Saldana@sms.ed.ac.uk), Kenny Smith, Simon Kirby & Jennifer Culbertson

Centre for Language Evolution, School of Philosophy, Psychology and Language Sciences,
The University of Edinburgh, Dugald Stewart Building, 3 Charles Street, Edinburgh, EH8 9AD, UK

Abstract

Human languages contain very little unconditioned variation. In contexts where language learners are exposed to input that contains inconsistencies, they tend to regularise it, either by eliminating competing variants, or conditioning variant use on the context. In the present study we compare regularisation behaviour across linguistic levels, looking at how adult learners respond to variability in morphology and word order. Our results suggest similar strengths in regularisation between linguistic levels given input languages whose complexity is comparable.

Keywords: artificial language learning; statistical learning; regularisation; variation; complexity; morphology; word order

Introduction

While languages exhibit variation at all linguistic levels, in the form of paraphrases, synonyms, allomorphs and allophones, that variation tends to be predictable: the choice of variant is (at least partially) conditioned by some aspect of the social or linguistic context. Occasionally, language learners are exposed to input that involves inconsistencies, for instance, when new variants are introduced into an established system, or when conventions are still not established, as in emerging languages (Senghas & Coppola, 2001; Siegel, 2004). Learners under those circumstances tend to reduce or remove such inconsistencies, i.e. they regularise their input. This can be achieved either by removing competing variants, or conditioning variant choice on the context (Ferdinand, Kirby, & Smith, 2017).

Regularisation has been documented extensively across linguistic levels (i.e. phonology, morphology, syntax and the lexicon) in natural language; e.g. in language acquisition, language change, and in emerging languages (Senghas & Coppola, 2001; Siegel, 2004; van Trijp, 2013). Experimental studies involving artificial language learning and statistical learning techniques report regularisation behaviour during the learning and production of probabilistic unconditioned variation in different linguistic units, across different linguistic levels (Culbertson, Smolensky, & Legendre, 2012; Fehér, Wonnacott, & Smith, 2016; Hudson Kam & Newport, 2005, 2009; Wonnacott & Newport, 2005). Nevertheless, it still remains an open question whether regularisation behaviour applies with uniform strength across linguistic levels and to what extent level-specific biases interact with regularisation during language learning and use.

Level-specific effects in regularisation behaviour

Research in second language acquisition and pidgin and creole studies has highlighted different developmental paths for morphology and syntax cross-linguistically (Good, 2015; Slabakova, 2013). Studies in pidginisation suggest that, in

periods when pidgins are highly inconsistent, linguistic levels might behave differently: morphologically complex traits such as inflectional morphology seem to be highly simplified whilst syntactic traits such as word order tend to reproduce the input complexity more closely (Good, 2015; Siegel, 2004). Good (2015) argues that this asymmetry is given by a break in transmission from source languages for morphological traits, which are only successfully transmitted if an entire contrasting paradigm is available to the learner, which is not the case in periods of linguistic instability. However, word order variation can be contrastive as well (e.g. S-Aux inversion to distinguish illocutionary forces). Alternatively, a more parsimonious hypothesis we could entertain is that a general tendency for pidgins to comprise highly simplified morphological traits and more conservative word order is rooted in the differing complexity of these traits in the source languages; Hudson Kam and Newport (2009) show that learners are more likely to regularise complex systems of variation.

Recent experimental studies have separately explored the effect of learning biases on typological asymmetries found in morphology and word order respectively. In morphology for example, St Clair, Monaghan, and Ramscar (2009) provide evidence of a preference for suffixing over prefixing, mirroring the cross-linguistic preference for suffixing. In word order, Culbertson et al. (2012) show that learners prefer consistent harmonic word order patterns (i.e. all modifiers either pre-nominal or post-nominal), also found more commonly in the world's languages. Moreover, Culbertson et al. (2012) show that this bias leads to different regularisation behaviour for different word order patterns. Nevertheless, no study has hitherto tried to systematically compare regularisation behaviour *across* linguistic levels. Uncovering differences in regularisation behaviour across linguistic levels could shed light on the intriguing asymmetry found in pidgin languages: morphological paradigms seem to be highly simplified whilst input complexity is more closely reproduced in word order.

In the present study we combine artificial language learning and statistical learning techniques to systematically compare the strength of regularisation of inflectional morphology and word order, controlling for asymmetries in the complexity and variability of the input languages.

Experiment 1

We utilise the methodology developed in Culbertson et al. (2012); Hudson Kam and Newport (2005). Adult learners are exposed to a miniature artificial language featuring an inconsistent mixture of synonymous variants. We are interested in how learners restructure the probabilistic unconditioned variation in the input languages, and to what extent that

restructuring is comparable across linguistic levels (specifically, morphology and word order).

Method

Participants Fifty-six native-English speakers (aged between 18 and 41, $mean = 23.2$) were recruited from the University of Edinburgh’s Careers Service database of vacancies. Each was compensated £6. Twenty-six participants were assigned to the Morphology condition, and 26 to the Word Order condition; the data from a further 4 participants (all in the morphology condition) were excluded as they either failed to learn the noun lexicon or failed to learn the associations between phrases and pictures.

Input languages We designed two novel languages which contained probabilistic unconditioned variation either in morphology or word order. Their respective probabilistic grammars are shown in Table 1. Both languages were used to describe simple pictures featuring one of two objects. Each object appeared either singly or in a pair; and could appear either in greyscale or coloured in blue. Descriptions were noun phrases composed of a Noun plus a Num(eral) and/or Adj(ective) modifier, which were presented orthographically and aurally to participants during the experiment.

All lexical items were 5 graphemes/phonemes long and had a neighbourhood density of 0 in the English lexicon. Nouns and modifiers differed in their syllabic structure; while all were bisyllabic, nouns (i.e. “mokte” and “jelpa”) conformed to a CVC.CV pattern, and modifiers to CV.CCV (based on English phonotactics and the Maximal Onset Principle).

Procedure Participants worked through a six-stage training and testing regime.

Stage 1, noun familiarisation Participants were trained on the two bare nouns that corresponded to pictures of the two different objects in the artificial language. During this phase, participants underwent a block of training consisting of 6 exposure trials and 4 picture-selection comprehension trials (in that order) —each noun-picture pair appeared 5 times (order randomised). Common to all training blocks to follow, on each exposure trial participants were presented with a picture (in this block always of a single object in grey-scale) and a corresponding description in the language (in this block, a bare noun), displayed both visually and aurally. On comprehension trials, participants were asked to select a picture out of an array of four (in this stage, the two objects seen during training plus two distractors) that corresponded to the displayed description in the alien language, and received feedback on their accuracy.

Stage 2, one-modifier training In Stage 2 participants were trained on one-modifier NPs, i.e. a Noun plus either Num or Adj only. Pictures contained any of the two objects presented either in blue and singly (Adj only) or in greyscale and in pairs (Num only). For each picture, a variant was selected randomly from the grammar assigned to the participant. Both grammars contained majority variants with an

Table 1: Probabilistic input languages in the Morphology and Word order conditions. Languages contain probabilistic unconditioned variation in inflectional morphology or word order respectively. All morphological variation resides in the suffixation of the modifiers. All word order variants conform to constituent structure [Num [Adj N]]. There are three types of NPs: Num Only (single Num modifier) refer to objects in pairs and in grey-scale, Adj Only (single Adj modifier) refer to a single object coloured in blue, and two-Mod(ifier) NPs (with both Num and Adj modifiers) correspond to objects in pairs coloured in blue. Languages include two different nouns (each corresponding to a different object) and thus comprise a total of 16 NPs (8 per noun) that correspond to a total of 6 pictures (1 per NP type, 3 per object).

NP TYPE	MORPHOLOGY CONDITION	WORD ORDER CONDITION
NUM ONLY	0.6 NP → N nefri	0.6 NP → N nefri
	0.4 NP → N nezno	0.4 NP → nefri N
ADJ ONLY	0.6 NP → N kogla	0.6 NP → N kogla
	0.4 NP → N kospu	0.4 NP → kogla N
TWO MOD	0.6 NP → N kogla nefri	0.6 NP → N kogla nefri
	0.13 NP → N kogla nezno	0.13 NP → nefri kogla N
	0.13 NP → N kospu nefri	0.13 NP → nefri N kogla
	0.13 NP → N kospu nezno	0.13 NP → kogla N nefri

empirical probability of $P = 0.6$, and minority variants with $P = 0.4$, as shown in Table 1. This phase comprised 40 trials in total, divided in 2 blocks of 20 trials; each block consisted of 15 exposure trials followed by 5 picture-selection trials. Participants saw each of the four different one-modifier pictures 5 times per block (order randomised).

Stage 3, one-modifier testing Stage 3 of the experiment tested the participants’ knowledge of the language. They saw the same pictures used in Stage 2 without accompanying text or audio and were asked to type in an appropriate description. They had to describe 20 pictures in total; each of the four different one-modifier pictures was presented 5 times in random order.

Stage 4, full training In Stage 4 participants were trained on a mix of one-modifier (a noun plus Adj or Num) and two-modifier NPs (a noun plus both Num and Adj). Two-modifier NPs were used to describe pairs of blue objects. For one-modifier phrases, variants were chosen in the same way as in Stage 2. For two-modifier phrases, variants were also selected randomly from the grammars assigned, with empirical probabilities of $P = 0.6$ and $P = 0.13$ for the majority and the three minority variants respectively (see Table 1). This stage comprised 100 trials (20 Num Only, 20 Adj Only and 60 two-Mod), divided into 4 block of 25 (15 exposure train-

ing trials followed by 10 picture-selection trials). Participants saw each of the four one-modifier pictures 10 times, and each of the two two-modifier pictures 30 times.

Stage 5, full testing Stage 5 tested participants’ knowledge of the whole language. They saw all pictures they had been trained on and were asked to type in appropriate descriptions. They had to describe 52 pictures in total: 10 Adj Only (5 per object), 10 Num Only (5 per object), 30 two-modifier (15 per object), and additionally, 2 pictures of bare objects by themselves and in grey-scale (1 per object).

Results

Output variability Figure 1 shows the entropy of participants’ production systems for both the Morphology and Word Order conditions. Analyses are run on Stage 5’s testing exclusively, i.e. participants’ final production sets. Words in the productions were corrected for typos (and only typos). Shannon entropy measures how variable participants’ productions are; the higher the scores, the more variable and the lower the scores, the more regular. The Shannon entropy (H) of phrase use for participant is given by

$$H(X) = - \sum_{i=1}^n P(x_i) \log_2 P(x_i) \quad (1)$$

where the sum is over the different variants, and $P(x_i)$ is the empirical probability of variant x_i in the set of a participant’s productions, X . We treated the two nouns for the different objects as the same variant when we calculated the entropy of the phrase variants such that no variability is introduced by the correct use of the different nouns. Entropy lower- and upper- bounds will vary depending on the number of required and possible variants as well as on the number of production trials. The most regular expressive language contains only one-to-one picture-phrase mappings and therefore only three different variants, one Num Only (e.g. *N nefri*), one Adj Only (e.g. *N kogla*) and one two-modifier (e.g. *N kogla nefri*). The final production phase consisted of 50 trials (excluding the two bare noun trials), divided up into 20 one-modifier trials (half Num Only and half Adj Only) and 30 two-modifier trials: the entropy lower bound for the language *overall* is thus 1.37 bits, and 0 bits for each of the NP types.

Figure 1 shows the entropy scores for the set of all participants’ productions (i.e. the overall language), as well as those for the production sets for specific NP types in isolation: one-modifier Num (Num Only), one-modifier Adj (Adj Only), and two-modifier (two-Mod) NPs. Entropy lower bounds and input entropies are represented as solid and dotted vertical lines respectively. A visual inspection of the Morphology and Word Order conditions in Figure 1 suggests that in many cases participants failed to reproduce the full variability of the input languages; entropy scores are generally lower.

We used the *stats* and *lme4* packages developed in R (Bates, Mächler, Bolker, & Walker, 2015; R Core Team, 2015) to run a linear mixed effects regression model (which we will call Model 1) to explore the effect of condition on

Table 2: Central tendencies of the proportion of majority input variants in production by condition and NP type. From left to right, the mean, median and mode(s).

		Proportion Majority Input Variant in Production		
		<i>mean</i>	<i>median</i>	<i>mode(s)</i>
Morphology	Num Only	0.704	0.8	0.919
	Adj Only	0.669	0.7	0.916
	two-Mod	0.609	0.65	0.843
Word Order	Num Only	0.580	0.65	0.094 & 0.96
	Adj Only	0.585	0.7	0.104 & 0.947
	two-Mod	0.442	0.33	0.089 & 0.92

regularisation behaviour (dependent variable: entropy). As fixed effects we entered Condition (two levels: Morphology as reference, and Word Order), NP Type (reverse Helmert coded with the 3 ordered levels: Num Only, Adj Only and two-Mod) and System (two levels: Input as reference, and Output). We also entered all interactions between fixed effects. As random effects, we included intercepts for Subject as well as by-Subject slopes for the effects of NP Type and System type. P-values were obtained through the *lmerTest* package (Kuznetsova, Bruun Brockhoff, & Haubo Bojesen Christensen, 2015). Results show a significant effect of System ($\beta = -0.346$, $SE = 0.085$, $p < .001$), suggesting that participants did indeed regularise their input in their output productions. We also found a significant interaction between System and Condition ($\beta = -0.284$, $SE = 0.119$, $p = .021$), suggesting that participants regularised their input significantly more in the Word Order condition. Results show the expected effect of higher input entropies in two-Mod NPs ($\beta = 0.21$, $SE = 0.024$, $p < .001$), and no significant interactions between NP Type and System (largest: $\beta = 0.027$, $SE = 0.028$, $p = .324$) or between NP Type, System and Condition (largest: $\beta = -0.041$, $SE = 0.039$, $p = .299$). These results suggest that participants regularised their input systems across conditions and NP types, and that participants in the Word Order condition regularised them more than those in the Morphology condition.

Variant production Table 2 provides the central tendencies for proportion use of the majority input variant for each NP type. We observe that all distributions in the Word Order condition are bimodal, with modes of the distributions of majority variant use at $P \leq 0.1$ and $P > 0.9$ across NP types, suggesting two opposite trends amongst participants: one towards the over-production of the majority input word order variants and another, towards their under-production.

Participants under-producing the majority word order variant in one-modifier NPs are necessarily producing modifiers pre-nominally. Figure 2 shows the overall proportions of the variants produced for two-Mod NPs by all participants. The input proportions are represented by the yellow vertical lines. The word order produced the most is the majority input variant N Adj Num. Although the three remaining input variants (below the grey solid line division) were equally frequent in the input language, the Num Adj N word order is overall

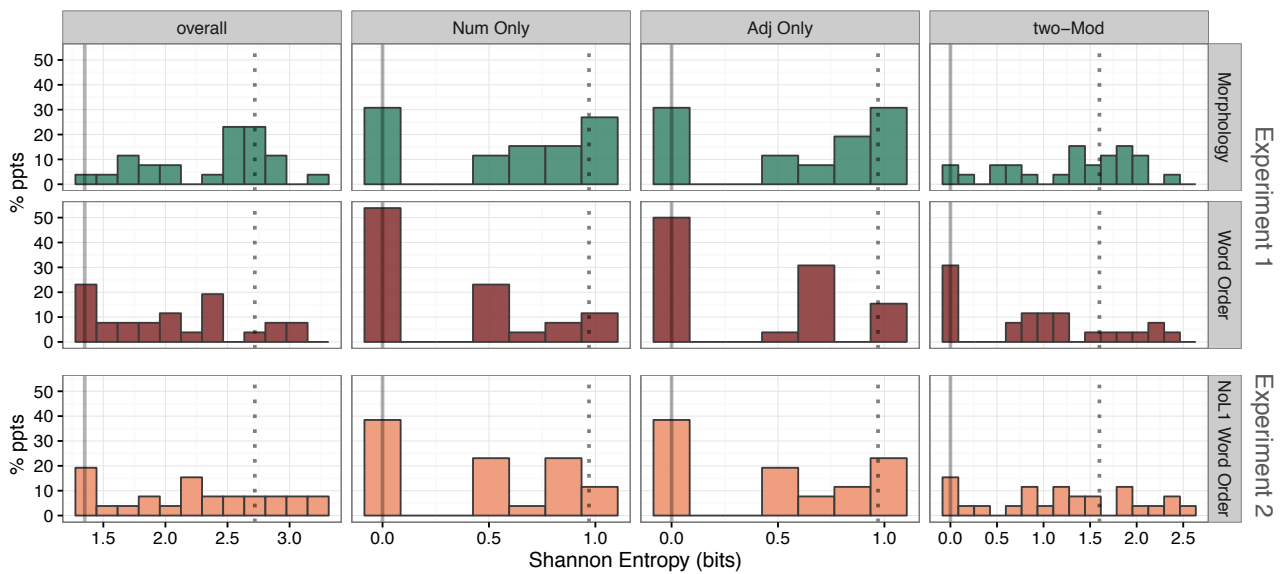


Figure 1: Entropy scores of participants’ production systems. From top to bottom, scores for the Morphology (green) and Word Order (red) conditions in Experiment 1 and for the NoL1 Word Order condition (orange) in Experiment 2. From left to right, entropies of participants’ full production sets as well as entropies by NP type: one-modifier Num (Num Only), one-modifier Adj (Adj Only), and two-modifier (two-Mod) NPs. Input entropy scores are indicated by dashed vertical lines. Minimum entropy scores are indicated by solid vertical lines. Minimum entropy is always 0 for each NP type in isolation but 1.37 for the overall system as it necessitates a minimum of 3 variants, one per NP type.

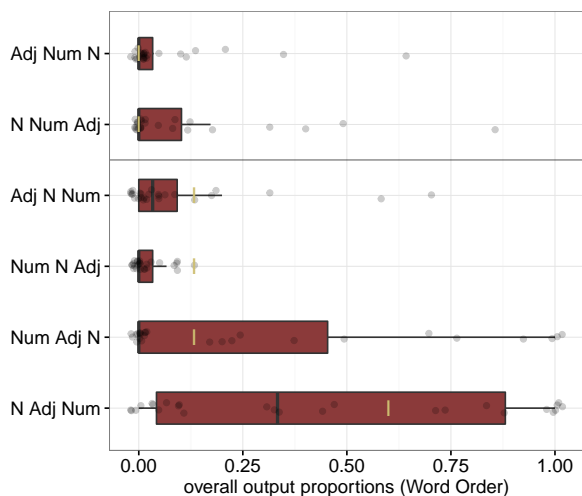


Figure 2: Box plot displaying the output proportions of two-modifier variants in the Word Order condition with individual participants’ data points overlaid. Seen (bottom) and unseen (top) variants during training are divided by a solid grey line. Vertical yellow lines indicate input proportions.

more frequently used (although only by a minority as indicated by the median value 0). Only 30% of participants produced systems with both harmonic variants (Num Adj N and N Adj Num) —and only 19% produced both variants more than once, suggesting that although both harmonic orders are preferred overall, they do not generally coexist within the productions of a single participant.

We ran a logistic regression model, which we will call Model 2, to explore the average difference between the proportions of Num Adj N variants in input and output linguistic systems. We entered System (two levels: Input as reference, and Output) as the only fixed effect. Random intercepts for Subject as well as by-Subject random slopes for the effect of System were also included. Results show that the Num Adj N variant is produced significantly less in output languages than in the input language ($\beta = -7.641$, $SE = 1.943$, $p < .001$). Only a minority of participants overproduced this variant, the majority of participants were in fact under-producing it. On top of the observed preference for harmonic order, these results confirm a tendency to avoid systems with two opposite N-peripheral variants, i.e. N Adj Num and Num Adj N.

Discussion of Experiment 1

Our results provide evidence that learners regularise probabilistic unconditioned variation in both morphology and word order. Regularisation behaviour is in line with an overarching simplicity bias argued to be at play in language learning and use (Culbertson & Kirby, 2016). Though the input languages were similar in terms of overall system complexity, regularisation behaviour was slightly stronger in the Word Order condition than in the Morphology condition. A close analysis of the variant usage in the Word Order condition suggests that this difference is driven by a bias in favour of harmonic N Adj Num and Num Adj N variants but against their coexistence within a system. This bias could be the result of L1 transfer; participants may have overproduced the Num Adj

Table 3: Probabilistic input language in the NoL1 Word order condition in contrast to the Word Order condition in Experiment 1. Changes in the variant set are indicated with boxes.

NP TYPE	WORD ORDER	NO L1 WORD ORDER
NUM ONLY	0.6 NP → N nefri	0.6 NP → N nefri
	0.4 NP → nefri N	0.4 NP → nefri N
ADJ ONLY	0.6 NP → N kogla	0.6 NP → N kogla
	0.4 NP → kogla N	0.4 NP → kogla N
TWO MOD	0.6 NP → N kogla nefri	0.6 NP → N kogla nefri
	0.13 NP → nefri kogla N	0.13 NP → N nefri kogla
	0.13 NP → nefri N kogla	0.13 NP → nefri N kogla
	0.13 NP → kogla N nefri	0.13 NP → kogla N nefri

N word order because it is the most common order in their L1 grammar. To minimise the possible effects of this level-specific word order bias, Experiment 2 investigated learning in a second word order condition, removing the English-like two-modifier harmonic pattern from the input.

Experiment 2

Experiment 2 follows the same design as the Word Order condition described in Experiment 1, with one difference: the set of two-modifier NP input variants. As illustrated in Table 3, we replaced the Num Adj N variant with the N Num Adj pattern, maintaining the number of harmonic word orders (two, i.e. N Adj Num and N Num Adj) but eliminating the L1 variant and the presence of opposite N-peripheral patterns. For ease of reference, we call Experiment 2 the NoL1 Word Order condition. We expect the change in the input language to mitigate the effect of L1 transfer and to increase the coexistence of both harmonic patterns.

Participants Twenty-eight native-English speakers (aged between 18 and 35, $mean = 24.8$) were recruited via the University of Edinburgh’s Careers Service advertisement database. Participants received £6. Only the data from 26 participants were fit for analysis as two participants either failed to learn the noun lexicon or failed to learn the associations between phrases and pictures.

Results

Entropy scores obtained in the NoL1 Word Order condition are shown in Figure 1 (coloured in orange). We ran a linear mixed effects model as in Experiment 1 to explore the effect of condition on regularisation behaviour (dependent variable: entropy), including the conditions in Experiment 1 plus NoL1 Word Order. The mixed-effects structure was the same as in the Model 1 but with reverse Helmert coding of Condi-

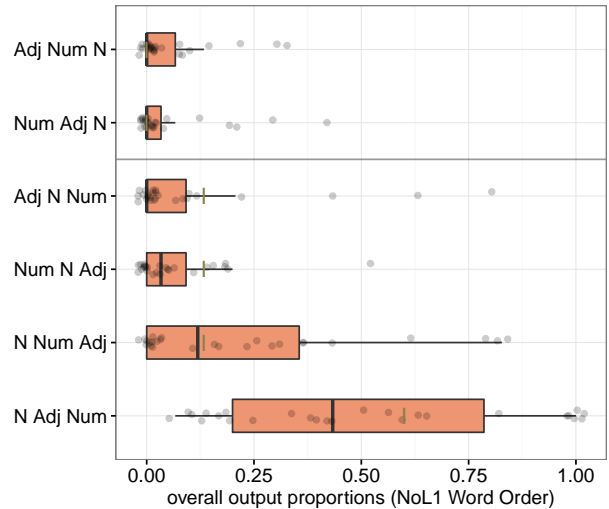


Figure 3: Box plot displaying the output proportions of two-modifier variants in the NoL1 Word Order condition with individual participants’ data points overlaid. Divided by a solid grey line, seen (bottom) and unseen (top) variants during training. Vertical light brown lines indicate input proportions.

tion such that NoL1 Word Order was directly compared to the Morphology condition from Experiment 1, and the Word Order condition was compared to the average of the Morphology and NoL1 Word Order conditions. Results show a significant effect of System ($\beta = -0.483$, $SE = 0.051$, $p < .001$) and a significant interaction between Word Order and System ($\beta = -0.073$, $SE = 0.036$, $p = .046$), ratifying the results in Model 1. However, we did not find a significant interaction between NoL1 Word Order and System ($\beta = -0.063$, $SE = 0.063$, $p = .317$), suggesting that participants in the Morphology and the NoL1 Word Order conditions regularised their input to similar degrees, and on average they regularised it less than participants in the Word Order condition in Experiment 1. As in Model 1, we did not find significant interactions between NP Type and System (largest: $\beta = 0.016$, $SE = 0.015$, $p = .288$) or between NP Type, System and Condition (largest: $\beta = -0.015$, $SE = 0.011$, $p = .168$). These results suggest that participants regularised their input systems across conditions and NP types, and that whilst participants in the Word Order condition regularised more than those in the Morphology condition, participants in the Morphology and the NoL1 Word Order conditions regularised their input to similar degrees. Excluding the Num Adj N variant in the input language thus eliminated the difference between levels. In other words, participants do not regularise probabilistic unconditioned variation in word order more than in morphology.

Figure 3 shows the overall proportions of the variants produced for two-Mod NPs in the NoL1 Word Order condition. We observe that the most produced word order is the majority input variant N Adj Num, and that the harmonic N Num

Adj word order is overall more frequent than any other minority input variant. Unlike in the Word Order condition where systems with both Num Adj N and N Adj Num patterns were not common, 65% of participants produced systems with both N Adj Num and N Num Adj harmonic variants in the NoL1 Word Order condition. We ran a logistic regression model to test the difference between the proportions of N Num Adj variants in input and output linguistic systems across participants. We used the same mixed-effects structure as in Model 2. Results suggest that the proportion of N Num Adj variants in the output languages is not significantly different from the input proportion across participants ($\beta = -0.594$, $SE = 0.546$, $p = .277$).

Discussion

Our experimental results reveal regularisation behaviour in the production of complex systems of variation in morphology and word order. They also suggest that regularisation behaviour is of similar strength between these linguistic levels given input languages with comparable initial complexities. In Experiment 1 we found higher levels of regularisation in word order than in morphology, apparently due to the specific properties of the set of variants in the input languages. When both harmonic pre-nominal and post-nominal two-modifier variants were included, the coexistence of both variants in a single production system was rare. Although a preference for harmonic order and consistent head position may have been at play, the interference of L1 transfer cannot be categorically rejected. Indeed previous research suggests that L2 learners tend to access their L1 knowledge if it matches the novel input (Weber, Christiansen, Petersson, Indefrey, & Hagoort, 2016). In Experiment 2, we showed that eliminating opposite N-peripheral positions in the subset of two-modifier variants by replacing Num Adj N with N Num Adj eliminates the difference in regularisation between levels. Our results do not suggest general level-specific learning biases that could straightforwardly predict a typological asymmetry between the strength and speed of regularisation in morphology and word order hinted at in pidgin and creole studies (Good, 2015). Instead, they suggest that asymmetries in regularisation processes in language formation ought to be sought in asymmetries in the input complexity of traits across levels, also taking into account the overlap of features between contributing languages.

Conclusion

Our results suggest similar strengths of regularisation between linguistic levels given input languages with comparable initial complexities. Nevertheless, preferences for certain patterns within a linguistic level might in fact vary the strength of regularisation behaviour within a given level.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.

- Culbertson, J., & Kirby, S. (2016). Simplicity and specificity in language: Domain-general biases have domain-specific effects. *Frontiers in Psychology*, 6, 1964.
- Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, 122(3), 306–329.
- Fehér, O., Wonnacott, E., & Smith, K. (2016). Structural priming in artificial languages and the regularisation of unpredictable variation. *Journal of Memory and Language*, 91, 158–180.
- Ferdinand, V., Kirby, S., & Smith, K. (2017). The cognitive roots of regularization in language. *arXiv preprint arXiv:1703.03442*.
- Good, J. (2015). Paradigmatic complexity in pidgins and creoles. *Word Structure*, 8(2), 184–227.
- Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language learning and development*, 1(2), 151–195.
- Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive psychology*, 59(1), 30–66.
- Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2015). lmerTest: Tests in linear mixed effects models [Computer software manual]. Retrieved from <http://CRAN.R-project.org/package=lmerTest>
- R Core Team. (2015). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Senghas, A., & Coppola, M. (2001). Children creating language: How Nicaraguan sign language acquired a spatial grammar. *Psychological science*, 12(4), 323–328.
- Siegel, J. (2004). Morphological simplicity in pidgins and creoles. *Journal of Pidgin and Creole Languages*, 19(1), 139–162.
- Slabakova, R. (2013). What is easy and what is hard to acquire in a second language. *Contemporary approaches to second language acquisition*, 9, 5.
- St Clair, M. C., Monaghan, P., & Ramscar, M. (2009). Relationships between language structure and language learning: The suffixing preference and grammatical categorization. *Cognitive Science*, 33(7), 1317–1329.
- van Trijp, R. (2013). Linguistic assessment criteria for explaining language change: A case study on syncretism in German definite articles. *Language Dynamics and Change*, 3(1), 105–132.
- Weber, K., Christiansen, M. H., Petersson, K. M., Indefrey, P., & Hagoort, P. (2016). fMRI syntactic and lexical repetition effects reveal the initial stages of learning a new language. *Journal of Neuroscience*, 36(26), 6872–6880.
- Wonnacott, E., & Newport, E. L. (2005). Novelty and regularization: The effect of novel instances on rule formation. In *BuclD 29: Proceedings of the 29th annual Boston university conference on language development* (pp. 663–673).

Appendix B

Experimental data

The data sets from the experiments discussed in this thesis are uploaded on DataShare and can be retrieved from <http://dx.doi.org/10.7488/ds/2235> (Saldana 2017).

B.1 Chapter 2

The workbook ‘Chapter 2 Experimental Data’ contains all the languages obtained in experiments 1, 2 and 3 in Chapter 2. The corresponding README file provides the keys to understanding the data file. Each spreadsheet within the workbook contains the data from a single transmission chain: the initial language and all the languages produced at each generation. There are four transmission chains per experiment and thus 12 sheets in total. Each spreadsheet is marked with a different colour that indicates the experiment they belong to: red for Experiment 1 (Artificial Only, chains A1-4; see section 2.2), blue for Exp 2 (Communication Only, chains C1-4; see section 2.3), and orange for Exp 3 (Communication + Artificial, chains CA1-4; see section 2.4). The number of the chains within experiment corresponds to the order they appear in the spreadsheet; e.g, the first spreadsheet within Exp 1 corresponds to chain A1 and the last spreadsheet of Exp 1 corresponds to chain A4.

B.2 Chapter 3

The data set for Experiment 4 in Chapter 3 is contained in the .csv file ‘Chapter 3 Experimental Data’. The corresponding README file provides the keys to understanding the data file.

B.3 Chapter 4

B.3.1 Experimental responses: languages and grammaticality judgments

All responses provided by participant during the experiments 5, 6 and 7 in Chapter 4 are contained in the .csv data file ‘Chapter 4 Experimental Data’. The corresponding README file describes the column heading of the data file. Note that the responses used to measure output variability are exclusively those in the final testing phase (see section 4.3.1.3), which is coded as ‘testing_all’ in the data set. The grammaticality judgments (see section 4.7.1.1) can be found under the ‘GrammaticalityJudgment’ phase.

B.3.2 Post-experimental questionnaires

Responses to post-experimental questionnaires (see section 4.7.1.2) are contained in the .csv data file ‘Chapter 4 Questionnaire Data’. The corresponding README file provides the keys to understanding the data file.