The Complexity of a Linguistic System is Shaped by the Communicative Needs of its Speakers

Vlad Nedelcu

Master of Science

Cognitive Science

School of Informatics

University of Edinburgh

2020

Abstract

Human language boasts an impressive balance between learnability and expressive power. According to the cultural evolutionary account, this balance is a consequence of language being repeatedly transmitted from one generation of learners to the next, as the pressures arising from this process cause it to change and adapt over time by developing complex structure. In this thesis, we use computational modelling to explore theories that linguistic form is also influenced by socio-cultural factors. We propose an extension to the Iterated Bayesian Leaning framework, which replaces the classic Bayesian agents with pragmatic agents based on the Rational Speech Act framework, introduces context-sensitive communication, and provides a model of speaker that is uncertain about the structure of the environment in which communication takes place. Using our model, we iterate a laboratory experiment which claims that when a speaker has less contextual information available to exploit, they will use more systematic utterances. We generally find these results to hold and be further propagated over cultural time. Our simulation results also strengthen the hypotheses that the languages of the earliest language-using communities would have been significantly less structured than those of modern-day communities. The extensions that we make to the original framework also open up exciting prospects for further investigating the role that social factors have on the evolution of language complexity.

Acknowledgements

First and foremost, I want to thank my supervisor, Kenny Smith, for his invaluable guidance, constructive criticism, swift replies, and most of all for his continued, active support in spite of the unprecedented times that we found ourselves in for the whole duration of this work.

I am also grateful to my family for their ever-present encouragement throughout the entire academic year. Needless to say that my trip to Edinburgh would have been impossible without their support.

Last but not least, I thank my friends from all across the different time zones.

Declaration

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

(*Vlad Nedelcu*)

Table of Contents

Chapter 1

Introduction

One particular aspect that sets human language apart from all other communication systems known in the natural world is the remarkable flexibility that it offers - its users can convey a seemingly boundless range of ideas, feelings, and thoughts. In spite of the complexity that language developed in order to accommodate this flexibility, children can get a good grasp of it in their first few years of life. This relative ease of learning would suggest that there is a strong fit between language and the human brain. The cultural evolutionary account (e.g., Christiansen and Chater 2008, Smith 2006) proposes that this fit is a consequence of language being repeatedly transmitted from one generation of learners to the next, as the pressures arising from this process cause it to change and adapt over time. Specifically, language has to be simple to learn, but also expressive enough so that its users can communicate without there being too much uncertainty over the intended meaning of an utterance (Kirby et al. 2004). However, for the creation of such a system to be possible, interlocutors must coordinate with one another, negotiating linguistic conventions (Brennan and Clark 1996), while also ensuring that these conventions are appropriate given the available contextual information (Wray and Grace 2007).

Computational models of cultural evolution, including the Iterated Bayesian Learning model (e.g., Kirby 2001), have previously shown how a communication system must develop systematicity in order to survive over a large number of transmission episodes. However, these models make no attempt to explore how the structure of the physical world, or environment, may impact the type of linguistic structure that emerges. Meanwhile, theoretical and experimental results indicate that environmental factors do matter, and that the structure of languages emerging through cultural evolution depends on the particular distribution of events that is specific to the world

observed by learners (Perfors and Navarro 2014), or on the way in which linguistic communities operate and use language (Wray and Grace 2007). Furthermore, experiments with human participants found that there is a correlation between the amount of contextual information that a speaker can utilize to reduce referential uncertainty in a given situation and the degree of systematicity in their utterances: less information means more structure (Winters et al. 2018). To investigate the consequences of this effect over cultural time, we propose an extension to the Iterated Bayesian Leaning framework, which supports communication in context and introduces pragmatic agents based on the Rational Speech Act model (Goodman and Frank 2016).

Evaluating the simulation results showed that contextual predictability is indeed a good predictor for the level of structure that emerges through cultural transmission: generally, when interlocutors have less contextual information in common, more information needs to be encoded inside their utterances instead, if they are to successfully communicate, so more structured communication systems will evolve, as they provide the best balance between communicative power and ease of learning. Moreover, in contexts where a less systematic language would suffice to guarantee communicative success, speaker uncertainty over the partner's specific type of context will gradually push the language towards more structure than needed. Our model also predicts that the languages of communities with heterogeneous communicative needs evolve to have more systematicity than their counterparts with more homogeneous needs. This strengthens the hypotheses (Wray and Grace 2007) that the languages of the earliest language-using communities would have been significantly less structured than those of the modern-day communities.

The target thesis starts by outlining some of the theoretical motivation of this work (section 2.1), briefly reviewing the rationale behind seeing language as a culturally transmitted system, as well as the consequences of doing so. We continue by presenting the frameworks on which our modelling work is based - Iterated Bayesian Learning and Rational Speech Act - and placing our approach in the context of these (sections 2.2 - 2.3). We then summarize the artificial language learning experiment that broadly guided our computational work, justifying some of our modelling choices (section 2.4). The following section describes the setup and implementation of the model, then discusses the simulation results, as well as the conclusions that can be drawn from these (section 3). Finally, in section 4, we conclude by reiterating some of our findings, and presenting some directions for future work.

Chapter 2

Background and related work

2.1 Motivation

2.1.1 Cultural evolution of language

For decades, accounts of the origins of human language have largely revolved around the existence of specialized brain mechanisms for processing language, with some variations on how we came to possess these (see Yang et al. 2017 for an in depth review). Most influentially, Chomsky (e.g., Chomsky 1965, Chomsky 1980) has argued about the existence of a series of innate constraints on language structure, which are collectively referred to as Universal Grammar (UG). In contrast to these previous theories, Christiansen and Chater (2008) propose that it is not humans who adapted to develop language, but instead language that adapted to humans. They argue that the speed at which language changes, coupled with the geographical dispersal of the human species, would have made it very unlikely for biological adaption or other non-adaptionist genetic mechanisms to result in the emergence of the largely arbitrary principles that are said to make up UG.

They instead propose understanding languages through the lens of an analogy to biological organisms, as ”highly complex systems of interconnected constraints” (Christiansen and Chater 2008: page 490) that have been shaped by adaptive pressures from the human brain, primarily related to processing and learning. They believe that language-specific mechanisms are unlikely to have significantly shaped the complex aspects of language structure, instead placing the emphasis on the role of more domain-general cognitive mechanisms. Specifically, through repeated episodes of cultural transmission, languages have gradually adapted to take a form that fits extant

human biases, so that they would be more readily acquired, understood, produced and eventually propagated.

To understand how cultural transmission can lead to these adaptions, we have to consider how children acquire their native language(s): through linguistic data. That is, children are brought up in the environment of their linguistic community, so the data that they observe is a direct product of the internal linguistic representation of the individuals belonging to that specific community (Smith 2006). When these children become linguistically mature, they themselves will produce data to be observed by new learners. Over time, as this cultural system gets repeatedly transmitted from a generation to the next, it will undergo various changes that will help it survive, by better adapting in response to the pressures that it faces. That is to say, the system culturally evolves. Some argue that these pressures come not only from the biases of individual learners, but also from the qualities that are inherent to the medium of transmission: articulation errors, disfluencies, noise, social factors, the impossibility of transmitting an infinite set of data (Kirby et al. 2004), etc..

The process of cultural evolution thus offers a possible explanation for many of the specific patterns that are observed in language structure (i.e. language universals): they arise through the repeated cycle of language learning and language use. During transmission episodes, languages always have to pass through the medium of transmission, and then through the minds of learners. This entails that, languages (or aspects of languages) that are not adapted to the pressures emerging from this cycle are unlikely to last and be attested. As a consequence, language universals can be viewed as statistical regularities across the attested languages of the world (Christiansen and Chater 2008). Later (see section 2.2), we go on to describe how compositionality, one of the key design features of language, can emerge as a result of pressures exerted during cultural transmission.

2.1.2 Group dynamics can shape language structure

It has been argued that the social contexts in which languages are used provide important sources of pressures. Wray and Grace (2007) argue that the dynamics of the community in which a language emerges are strongly correlated with the specific form and structure that the language takes. They propose that many of the features that are now widespread through the languages of the world, including those deemed as universal, are only manifested for ”the duration of the particular social and cultural context

that has spawned them" (Wray and Grace 2007: 544). Since the hunter-gatherer communities in which language first emerged functioned very differently from modern-day communities, this would have deep implications on our appreciation of what exactly is to be considered fundamental to language.

According to Thurston et al. (1987), languages can fulfill two communicative functions: esoteric and exoteric. Esoteric, or intra-group communication, is used in interactions among individuals of the same group. This means that interlocutors assume a considerable amount of common ground when engaging in this type of communication, because of their shared environment, cultural practices, beliefs and overall group knowledge. On the other hand, exoteric, or inter-group communication, is reserved for interaction between members of different communities. In this case, interlocutors cannot consistently rely on common ground to make themselves understood, having to express their messages in a much more explicit and systematic manner that makes as few assumptions as possible about their communicative partners.

Communities that are rather homogeneous and inward-facing will have their language more adapted towards esoteric communication, making it harder for strangers to understand them and less likely for group members to engage with outsiders (Wray and Grace 2007). As a result, the language will be mainly learned by children in the community, so it will have more complex patterns and less systematicity. Conversely, communities that are more outward-facing are likely to use a system that is well adapted for inter-group communication. This system will likely employ more structure and systematicity, which will make it more learnable for outsiders, who as a result are more readily integrated into the group.

Wray and Grace argue that the first language users lived in relatively small, geographically concentrated communities, which allowed them to establish intimate connections and significant common ground. Because of this, the default setting of human language is very likely to have been inherently esoteric. As groups started to grow in size and interact with one another, the socio-cultural factors also started to change, causing a gradual shift from esotericity to exotericity. By the time languages became attested, rich literary traditions would have already emerged across their communities, indicating that these languages have further shifted towards exotericity, enough so to allow communication about things that are geographically and historically remote.

We will use our model to explore their predictions regarding the role of sociocultural influences on the evolution of linguistic structure. Specifically, we will show how a community's shift towards exotericity is associated with an increase in the sys-

tematicity of their language. This is an important contribution of our work, as there are currently no computational models of the theories presented by Wray and Grace.

2.2 Iterated Bayesian Learning

As outlined in the previous section, the account of language evolution by cultural transmission holds that many of the unique characteristics exhibited by human language are a direct result of the fact that languages are transmitted from one generation to the next through a repeated cycle of language production, observation and learning. As such, the learners' individual differences are equated with the genetic variations of biological evolution, as the driving force behind language evolution (Christiansen and Chater 2016).

The Iterated Learning Model (ILM) was first introduced by Kirby (2001) to explore the effects of the cultural transmission of linguistic behaviour over a large time frame. In ILM, agents of a generation observe some linguistic data, infer from this data an internal language representation that will dictate how utterances can be generated for particular meanings, and using this representation, produce a novel set of utterances to be transmitted to the next generation of agents. This process is then repeated over a large number of generations.

To better understand the influence that the differences introduced by individual agents can have on the languages that evolve through iterated learning, Griffiths and Kalish (2007) imagine that these agents learn using Bayesian inference: in inferring a language, their prior beliefs are combined with the conclusions that they draw from the observed utterances produced by other agents. This alteration allows for the biases of learners to be made explicit through the prior, resulting in a flexible framework that has enabled the exploration of various linguistic phenomena, including regularization and the evolution of frequency distributions (Reali and Griffiths 2009), the origins of word-order universals, (Culbertson et al. 2012) and the evolution of structure in language (e.g., Kirby et al. 2015, Kirby et al. 2007).

Using the Iterated Bayesian Learning model, Kirby et al. (2015) offer an account for the evolution of one fundamental feature of human language: compositional structure. Specifically, it is because of the combinatorial nature of modern languages that complex utterances can be built by recombining smaller units from a simpler, largely fixed set (morphemes or words). The meanings of these larger utterances can then be systematically interpreted from the meanings of their constituent subunits, due to

compositionality. Henceforth, when we talk about different levels of structure in languages, we will refer to different rates of compositionality and systematicity in expressing meanings (more details in section 3.2).

Kirby et al. (2015) show that structured systems can emerge from the interplay of two pressures. First, a pressure for compressibility comes from our natural preference for more compact mental representations, which translates to a bias for simpler and more structured languages. Furthermore, a language would in principle be represented as an infinite set of utterances, while the agent can realistically only be exposed to a subset of these. This creates a data "bottleneck" on the iterated learning process, which further amplifies the agents' bias for systematicity. However, a communication system that is fully compressed would not prove itself to be too useful, despite being perfectly systematic: all possible meanings would be conveyed with a single maximally ambiguous signal. We can thus speculate that a second pressure must be involved, one which calls for a lexicon that allows as many meanings as possible to be unambiguously expressed, and which is imposed precisely by the act of communication. For an agent abiding by the principle of Bayesian inference, the bias for simplicity would be encoded into the prior, while the pressure for expressiveness would arise from the observed behaviour of other agents, considering that their behaviour is shaped by communicative goals. In their model, these goals are directly enforced as agents chose what utterance to produce, by penalising ambiguous utterances, which are inexact, and would sometimes lead to unsuccessful communication with the interlocutors. The authors show that in a setting characterized by naive agents in each generation, and a first generation that observes an unstructured language, the two pressures will indeed push the agents towards introducing structure.

The model that we later introduce will build on this specific work, and will explore what happens when additional pressures coming from the structure of the environment are added into the mix. On the one hand, we investigate how the amount of common ground between interlocutors affects the complexity of the evolved language, if it does at all. On the other, we explore how languages adapt to support communication between multiple interlocutors (i.e. in groups), with potentially different communicative needs, and uncertainty about the needs of their partners. We will also not enforce communicative goals through a direct penalty on production, instead letting these goals arise from the cooperative principle (see below).

2.3 The Rational Speech Act framework

Flexibility is one of the most important and distinctive features of human language, as it enables us to express any conceivable thought in any particular situation. This feature is partially due to the almost infinite interpretability that a linguistic utterance can have depending on the context in which it is used, and this represents a focal point in the study of pragmatics (Grice 1975). As such, it is believed that language is likely to be significantly shaped by pragmatic processes (Barron and Schneider 2009), so these should not be overlooked when modelling language evolution.

The Rational Speech Act (RSA) framework, due to Goodman and Stuhlmüller (2013), offers a Bayesian account for the process of pragmatic reasoning, based on a Gricean (Grice 1975) view of pragmatics. This view is rooted in the cooperative principle: two conversational partners act cooperatively and design their utterances to be as informative as necessary for the purpose of their conversation. RSA has already been successfully applied to investigate a wide range of pragmatic phenomena, including grounded word learning (Goodman and Frank 2016), colour identification in context (Monroe et al. 2017), informativeness in question-answering games (Hawkins et al. 2015), or the choice of taxonomic levels of reference (Degen et al. 2020).

The framework builds on the concept of *common ground*: some information i is common ground to two agents if they both have information i, they are both aware of the fact that they both have information i, they both know that they are both aware of the fact that they both have information i, and so on to infinity. When two agents are communicating, they are transmitting knowledge that is not in their common ground, and is instead the result of each observing a different state of the world and forming a different set of beliefs. Through acting cooperatively, they attempt to formulate utterances that bring together the two distinct sets of beliefs. They achieve this by recursively reasoning about each other: the speaker designs its utterances by thinking about how the listener would interpret them using their set of beliefs, and the listener interprets the utterances by trying to figure out what the speaker could have meant, considering that they were trying to be helpful and informative.

More formally, the pragmatic speaker chooses an utterance u from a set of alternative signals, by considering how certain a literal listener would be about the intended state of the world w after seeing the utterance:

$$P_S(u|w) \propto exp(\alpha \, log \, P_{Lit}(w|u)) \tag{2.1}$$

where α represents the rate of rationality.

The literal listener forms their beliefs by probing only the validity of the literal meaning of the utterance, while also considering the prior $P(w)$:

$$P_{Lit}(w|u) = \mathcal{L}(u,w)\,P(w) \tag{2.2}$$

$$\mathcal{L}(u,w) = \begin{cases} 1, & \text{u is true in world state w} \\ 0, & \text{otherwise} \end{cases} \tag{2.3}$$

The actual pragmatic listener then infers the new state of the world, by reasoning about what the pragmatic speaker might have meant given what they said in order to abide by the cooperative principle:

$$P_L(w|u) \propto P_S(u|w)\,P(w) \tag{2.4}$$

While we adhere to only illustrating the basic form of RSA here, numerous variations exist that provide more sophisticated models of listeners and speakers (Hawkins et al. 2017), more efficient and cognitively plausible models of speakers (White et al. 2020), or an account for communication through a noisy channel (Bergen et al. 2016). Later, we will introduce our version of the framework, providing a model for a speaker that is uncertain about the state of the world in which communication takes place. This version will be used for the interactive aspect of our model, with interlocutors resolving ambiguities by directly reasoning about their communication partners when producing or interpreting linguistic utterances. This is in contrast to Kirby et al. (2015), where a penalty for ambiguous utterances is introduced through an additional parameter.

2.4 The impact of environmental factors on language

Through iterated learning, it has been illustrated how linguistic structure could emerge when simultaneous pressures are exerted from the process of individual learning, and from the communicative act itself, due to its interactive nature. But, there has recently been increasing interest in studying how the structure of the physical environment in which communication takes place impacts the emergence of systematicity. For instance, Müller et al. (2019) show that consistent access to the visual context during communication supports convention formation between interlocutors, leading to higher rates of communicative success. Meanwhile, Nölle et al. (2018) found that

higher systematicity in language evolves in dynamic environments where the communicative context rapidly expands, as well as in environments where absent referents have to be frequently communicated, as a counterbalance for the increased working memory load on the speaker. However, in this section, we are going to focus on the artificial language learning experiment of Winters et al. (2018), which explores how contextual information directly interacts with the two aforementioned pressures, in order to shape language autonomy. In this paper, context is defined due to Sperber and Wilson (1986) as the "mutual cognitive environment" in which linguistic utterances are interpreted, thus governing the distinctions that must be made in order to reduce uncertainty. They continue to describe that some contexts are naturally more predictable than others, and that this property will affect the amount of information that can be estimated by a speaker, to be then utilized in order to minimize the ambiguity of their utterances for the listener. In the experiment, this contextual predictability is simplified and determined by the setting of only three components: *amount of shared knowledge*, *immediate context* and *historical context*.

Starting from Clark's notion of common ground, they define the more quantifiable concept of *shared knowledge*, as the total information that interlocutors have in common and that is relevant to their interaction. The hypothesis is that speakers will produce more transparent utterances when they perceive the listener as less informed about a situation, since there is less common knowledge to rely on. Because of this, the speaker's language will use more self-contained utterances, whose interpretation is not dependent on that particular situation or context (i.e. autonomous utterances). Consequently, the communication system itself will be more autonomous overall. In their experiment, the speaker can either share the complete communicative context with the listener (shared context), or only the referent that it will have to convey (unshared context). Therefore, more autonomy is expected to emerge in the unshared context setting.

The communication act can be split into a series of instances, where there is some distinctive discriminative piece of information that ensures an utterance is correctly comprehended at a particular point in time. This is captured by the *immediate context*. To illustrate this, consider that in the experiment all referents have two semantic dimensions: shape and colour. If all the referents in the communicative context are differentiable solely by colour, then the only relevant information that the listener needs in order to identify a particular object is its colour. Instead, if the objects that the listener sees differ in both shape and colour, then both dimensions are needed for discrimination. Consequently, a more predictable setup is one where only one of the two

dimensions is sufficient for successful communication between the interlocutors.

As interlocutors engage in communication, a series of conventions are established between them that govern the utterances picked to describe specific situations. *Historical contexts* refer to these negotiated conventions, which interact in an interesting way with the immediate contexts. Specifically, Brennan and Clark (1996) found that interlocutors will prefer to abide by strong, pre-established conventions (if those exist), even when a simpler strategy would suffice for successful communication in the current immediate context. In the experimental setup, this would entail that if the earlier contexts encountered by the interlocutors demanded the encoding of both semantic dimensions, then when they are presented with a context where one dimension is sufficient, there is nevertheless a tendency for them to prefer sticking with the already established conventions, even if that means transmitting redundant information. As such, the immediate context can sometimes be overridden because of prior events. In a predictable setup, interlocutors should have to renegotiate conventions as little as possible throughout their interaction. For example, if the first three contexts only required the transmission of colour, a setup where the next context suddenly demands the communication of both dimensions (i.e. some referents are identical in their colour) would be less predictable than a setup where specifying the colour remains a good strategy.

Their experiment consists of multiple rounds of pairwise interactions, in which the speaker has to communicate specific referents to the listener, so that the latter can pick them out among a number of distractors. Of the setups described earlier, a hierarchy of contextual predictability emerges, in which *shared context* + *shape-different context* sits at the top, while *unshared context* + *mixed context* sits at the bottom. Their findings show that this hierarchy predicts the level of autonomy of the language that the participants settle on: when overall predictability is high, autonomy is instead low, as there is more contextual information that the speaker can rely on; as predictability decreases, signal autonomy increases.

While this experiment provided novel insight into the role than environment has in shaping language, the methodology that was used makes it difficult to be certain that speakers are actually actively monitoring the needs of listeners when designing their utterances, and that this was the factor that caused the observed results. We should note that most other work on this topic (cited earlier) has also centred around the artificial language learning framework. As such, one of the main contributions of our project consisted in computationally modelling the interaction between all these pressures, in hopes of achieving a better understanding of the phenomena involved.

Chapter 3

Description of work

In order to formalize the aforementioned experiment and explore the hypotheses of Winters et al. (2018) in more detail, we built an evolutionary agent-based model, consisting of language learning, language use and cultural transmission. In similar fashion to Kirby et al. (2015), learning is implemented as a process of Bayesian inference: the agents first observe some linguistic data, then form a hypothesis about how that data could have been generated. Language use is modelled as a problem of pragmatic reasoning, in accordance with the Rational Speech Act framework: agents acting as speakers are aiming to produce utterances that are informative to listeners relative to their particular needs, while agents acting as listeners are aware of this fact when interpreting said utterances. Specifically, the speaker attempts to convey a meaning to a listener, using an utterance that will be helpful in the particular communicative context that the listener finds itself in. Lastly, language transmission from one generation to the next is modelled following the iterated learning framework: the agents in a generation observe data produced by learners in the previous generation, while the data resulting from their own interactions will be observed by the following generation.

Throughout this section, these components will be presented in more detail: starting with the overall setup of the task, followed by the nature and representation of the languages and communicative contexts inferred by the agents, the form of the hypothesis space, as well as the prior over this space, and finally the complete formal description of the main models, coupled with the results. The last subsection presents an attempted alternative at working with a larger hypothesis space, one that did not however end up in the main models.

3.1 Overall setup

Following the Winters et al. experiment, the task is split into two main phases: the learning phase and the communication phase, with each phase being repeated for every new generation of agents (see figure 3.1.).

The learning phase takes place separately for each agent in a generation, and over multiple rounds per generation: in each round an agent is presented with a meaning-signal pair from a data set, the agent learns from the pair, then infers the languages that are most likely to have produced it. Since there is no interaction between agents in this phase, they are agnostic with respect to the context that the data pair could have resulted from. The source of the data set is either the previous generation's produced output during the communication phase, or a predefined set of pairs, in the case of the first generation. If the size of the data set is larger than the number of learning rounds, then the pairs used in this phase will be selected at random, otherwise the agents are prompted to learn at least once from each pair in the set.

During the communication phase, the agents are placed in an asymmetric communication game: in each round, the agents are assigned the role of either listener or speaker; the listeners, possibly multiple in number, are each provided with a target meaning that they will have to identify, as well as a number of distractors, which together form a communicative context (note that listeners do not know which meaning in the context is the target); the unique speaker, who knows the target, has the task of communicating to each listener a signal that will help it identify the correct referent within its specific context. After each round, the listeners receive feedback as to whether or not they correctly identified the intended meaning, and subsequently learn from this information before moving on to the next round. It is worth noting that this situation departs from the one modelled in Kirby et al. (2015) in that the size of the contexts is not necessarily maximal (i.e. composed of all the possibly conveyable meanings). One aspect would also seem to set our model apart from the experiment in Winters et al. (2018): learners can have their role switched from one communication round to the next (i.e. from listener to speaker or vice versa). In an experiment with human participants this would allow an individual acting as listener to easily figure out what type of context they were placed in, then directly use that information later on as a speaker, without having made use of the given feedback. However, we can ensure that our agents exclusively utilize the feedback to infer the context-type, thus still abiding by the conditions of the experiment. At the end of the generation, the signal-meaning

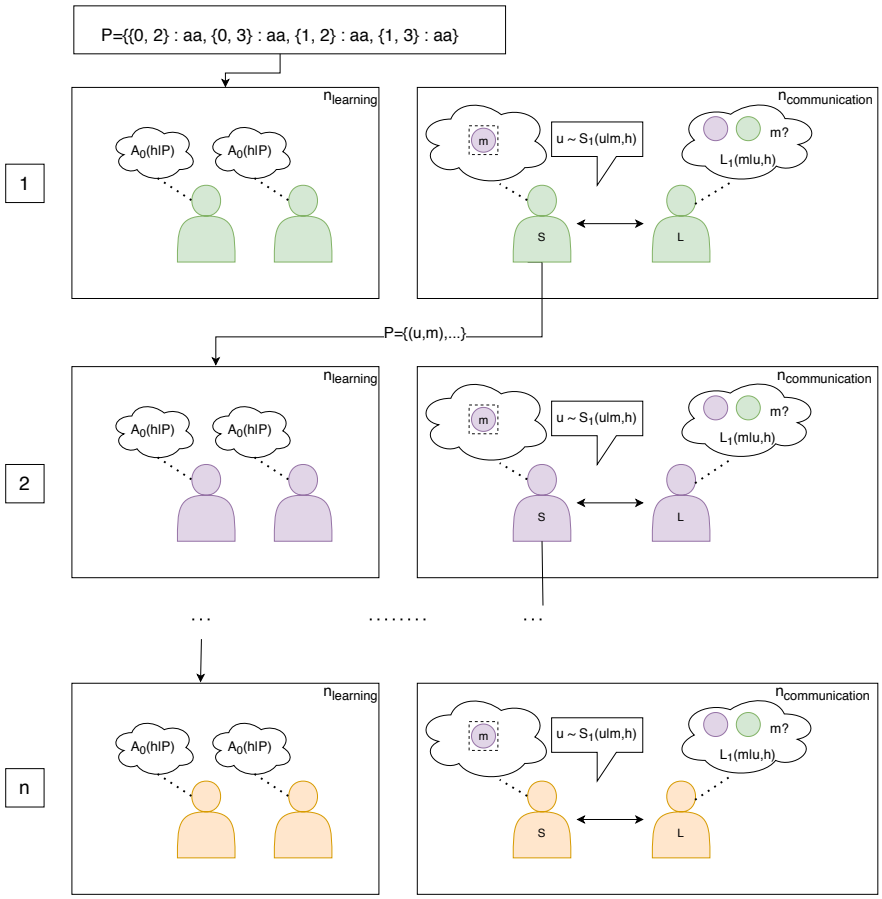

Figure 3.1: Simulation procedure following the Iterated Learning framework. n generations of distinct agents are arranged in the form of a chain. During the learning phase (left), agents update their posterior over the hypothesis space by learning from a set of meaning-signal pairs P, for a number of $n_{learning}$ rounds. During the communication phase (right), for $n_{communication}$ rounds, two agents are chosen to be placed in a communication game, one playing the role of speaker and the other of listener. The utterances produced by the speakers will be used as learning material for the next generation.

pairs produced during communication are used as learning material by the naive agents introduced in the next generation.

3.2 Languages

At the most elemental level, a language could be defined as a system that enables its users to convey meanings using signals. We represent languages as sets containing pairings of signals and their associated meanings. Our purpose is to study the emergence of structure in language, so this simple representation system must be capable of exhibiting systematicity. To achieve this, both meanings and signals are made up of smaller units: a meaning is a set of features that characterize the concept being referred, while a signal is an ordered sequence of characters. We chose minimal parameters for our experiments: meanings and signals of size two. For most of the models detailed in this section (with the exception of 3.10), the characters that form the signals are extracted from a vocabulary of size two: {a, b}, while the values of the first feature are extracted from {0, 1}, and of the second feature from {2, 3}. These choices are due to computational motivations, as they directly determine the size of the hypothesis space. In our case, the complete possible set of meanings is {aa, ab, ba, bb}, and that of signals is {{0, 2}, {0, 3}, {1, 2}, {1, 3}}, yielding 256 different languages.

These languages are grouped into classes depending on their level of systematicity: how consistent they are in their strategy of expressing meanings. The highest on this scale are degenerate languages, which express all meanings using a single, completely ambiguous signal (e.g. {{0, 2} : aa, {0, 3} : aa, {1, 2} : aa, {1, 3} : aa}). Next are one-feature-only languages, which map all meanings that are equal in one of the two features to the same signal (e.g. {{0, 2} : aa, {0, 3} : aa, {1, 2} : ba, {1, 3} : ba}), followed by compositional languages, which have consistent mappings for both features that make up the meaning (e.g. {{0, 2} : aa, {0, 3} : ab, {1, 2} : ba, {1, 3} : bb}). Hybrid languages, which have at least one ambiguous signal and mix the strategies used by the previous classes (e.g. {{0, 2} : aa, {0, 3} : aa, {1, 2} : ba, {1, 3} : bb}, sit on the scale just above holistic languages, which idiosyncratically map every meaning to a distinct signal (e.g. {{0, 2} : aa, {0, 3} : ab, {1, 2} : bb, {1, 3} : ba}), thus having the lowest level of systematicity. The motivation for this categories will be detailed in section 3.4.

3.3 Communicative contexts and their predictability

As in Winters et al. (2018), there are two main aspects that could differ in the setup of the game: the agents' access to the context (shared/unshared) and the type of the context (mixed/one-feature-different). For the first aspect, if access to the context is shared, than both the speaker and the listener are aware of the communicative context, whereas if it is unshared, the speaker only knows the meaning that it must convey to the listener, and not the distractors in its context. The second aspect, the context-type, determines the semantic features that will differ in contexts across the communication rounds. In a one-feature-different setup, only one feature is relevant across all the rounds for discriminating among the referents in a context, and that feature is furthermore consistent (i.e. only the first feature is ever different). This is not the case in a mixed setup, where the meanings of a context might differ in any of the two features. This yields that a more general strategy will need to be applied by the agents in the mixed setup, compared to the one-feature-different setup.

In our model, we represent a context as a set of meanings (e.g. {{0, 2}, {0, 3}, {1, 2}} is a context containing three meanings). A context-type is defined as the set of all contexts that are included in it, so {{{0, 2}, {0, 3}}, {{1, 2}, {1,3}}} would be a one-feature-different context-type (since only the second feature is sufficient to correctly discriminate between the meanings in both contexts), while {{{0, 2}, {0, 3}}, {{0, 2}, {1, 2}}, {{0, 3}, {1, 3}}} would be a mixed context-type (since the features that are the key to discriminating between the meanings of the context are not consistent).

These two aspects result in four different conditions, which can be observed in figure 3.2: shared access + one-feature-different context, shared access + mixed context, unshared access + one-feature different context, unshared access + mixed context. The first of these yields the highest contextual predictability, since the speaker knows that it is enough to communicate a single semantic feature to the listener in order to achieve successful communication, as that feature is consistently sufficient to distinguish between all the meanings in any context. In this case, we expect the agents to favour one-feature-only languages. On the other end of the spectrum, the fourth condition is the most unpredictable, with the speaker having to infer by itself that it has to communicate both features to the listener, and that any other strategy would not be guaranteed to succeed. Because of this, compositional and holistic languages would be the most suitable in this case. The second and third conditions are only partially predictable, so agents are expected to behave somewhere in between the more extreme conditions.

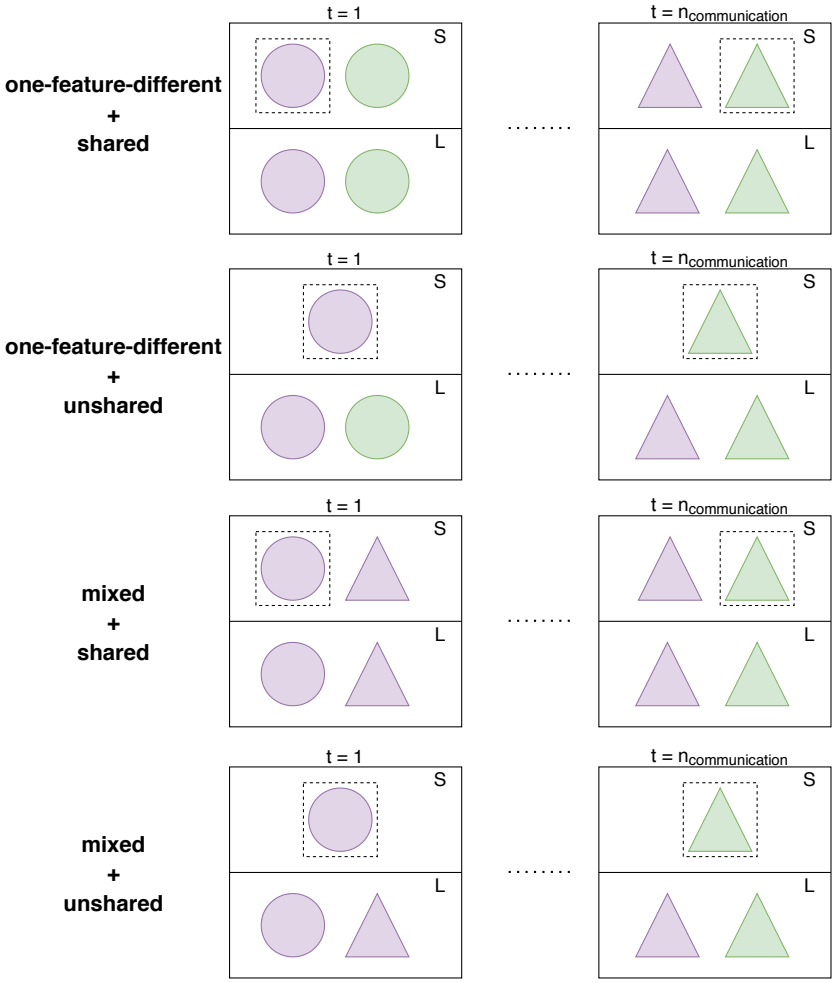

Figure 3.2: Examples of possible setups of the communication phase in each of the four conditions. The complete phase consists of $n_{communication}$ rounds, and each square represents the available contextual information from the perspective of the speaker (upper part) and listener (lower part) for one round. For representation purposes, we imagine that the referents are geometrical shapes, with shape and colour being the two characteristic features. In the one-feature-different + shared condition, both agents have access to the full context, and the referents are always distinguishable by colour only. The latter is true for the one-feature-different + unshared setup, but this time the speaker only sees the correct referent in isolation. In the two mixed setups at the bottom, referents sometimes differ in one shape (colour) and sometimes in the other (shape). The same distinction as earlier is made between shared and unshared variants.

3.4 Hypothesis space

The agents infer a distribution over tuples of languages and context-types, yielding that the hypothesis space is formed of all such possible distributions. Each hypothesis made by an agent is composed of a single language and a number of context-types equal to the number of distinct listener roles. The agents do not infer the forms of the context-types themselves (which are predefined), but binary variables that just discriminate between the mixed and one-feature-different context-types. Thus, the complete prior has two components: a prior over the language space and a prior over the context-type space. For our main model, we assume that at the start of the communication phase agents have no information about the context-type assigned to each listener role, so we use an agnostic, flat prior over the context-type space.

The prior over the language space was motivated by the scheme presented in Kirby et al. (2015): they determine for each language the associated construction grammar that generates it, then rewrite this grammar using a minimally redundant encoding; the prior is defined as a function of the total number of bits used for the encoding. In essence, this ends up favouring languages that are simple and systematic, as previously explained in this chapter.

However, for our purposes, using this exact scheme seemed to pose a series of problems. First of all, our task demands more granularity in the classification of languages, as Kirby et al. do not separate the set of one-feature-only languages from that of the more general "other" languages (what we call hybrid languages). Furthermore, this sets need to be clearly distinguishable from all other categories in terms of prior. However, compositional languages are given a prior that is very close to that of one-feature-only languages, making it more difficult for the agents to consistently determine the more compressible of the two types.

Furthermore, when communicating within mixed contexts, we found that agents would be prone to converge to a specific type of hybrid languages (e.g. $\{\{0, 2\}$: aa, $\{0, 3\}$: bb, $\{1, 2\}$: bb, $\{1, 3\}$: aa$\}$) instead of compositional ones. This is a consequence of the prior method attributing a higher prior to these hybrid languages when compared with compositional languages. The language above can also lead to successful communication within all contexts that differ in one feature (e.g. $\{\{0, 2\}, \{1, 2\}\}$). However, these languages seem rather unsystematic, as they have no consistent relations between signal units and meaning features. Because of their less predictable systematicity it could be argued that they should be attributed lower priors than com-

positional languages.

An additional problem arises as the sets of meanings and signals increase in size: the length of the encoding for degenerate languages increases significantly faster than that of compositional languages. This is because all individual meanings have to be separately specified in the rule for degenerate languages, while these meanings result from the composition of smaller rules for the compositional languages. Consequently, the language types will get closer and closer in terms of their prior, with compositional languages eventually being assigned a higher prior than degenerate languages for large enough meaning and signal sets.

For these reasons, we decided to manually set the priors, so that the hierarchy is still respected, but the language types are also clearly distinguishable for our purposes.

3.5 Main model

We now formally introduce the main computational model, restricted only to generations of size two (i.e. two agents per generation). Two different variants (with differences only in the speaker's communication phase) will be laid out, corresponding to the shared context setup and unshared context setup respectively.

It is worth mentioning that Bayesian models generally work with small numbers (as is our case), so rounding errors could be problematic. Even though we represent probabilities on a logarithmic scale in our actual implementation so as to mitigate potential issues, the models below are detailed in terms of classic probabilities for simplicity.

3.5.1 Setup

As previously described, the population is arranged in a chain structure: the naive agents of each generation are first trained on data produced by the previous generation, before starting to interact with each other in order to produce the data that will be forwarded to the next generation of agents. In each round of interactions an agent can be attributed one of two roles: speaker or listener. Throughout the whole simulation, the listener role will be associated with a single fixed context-type: one-feature-different or mixed.

3.5.1.1 Shared context

The two agents establish the communicative context c_l as a common ground: they share the same context and are both aware of this fact. This entails that the agent acting as speaker is aware on a round-by-round basis of the distinction that it has to make with its utterance in order to successfully convey the intended meaning to the listener.

3.5.1.2 Unshared context

In contrast to the previous case, the speaker sees the target referent that it must convey to the listener in isolation, without having access to the other distractors that are situated in the communicative context c_l. Therefore, this agent has no knowledge of what distinctions would be most useful on a round-by-round basis in order to help the listener identify the intended meaning. Instead, it will have to infer the context-type, then be informative on average with respect to the inferred type when designing its signal.

3.5.2 Learning phase

At the start of each generation, each agent a in the population updates its posterior distribution over the hypotheses A_a by learning from a set of pairings of signals and their associated meanings P. At the end of this phase, the posterior of a hypothesis h (composed of a language l and a context-type t), given the learning set P, is defined as:

$$A_a^0(h=(l,t)\,|\,P) \propto \prod_{(m,u)\in P} S_0(u\,|\,m,l)\,P(l)\,P(t) \tag{3.1}$$

$$S_0(u\,|\,m,l) = \begin{cases} \frac{1}{|p_m|} - \varepsilon, & (m,u) \in l; \\ \varepsilon_{err}, & (m,u) \notin l \end{cases} \tag{3.2}$$

where $|p_m|$ is the number of signals that map to the target meaning m in the language l; $P(l)$ and $P(t)$ are the priors over the language space and context-type space respectively; ε_{err} is the probability that the literal speaker makes a mistake and chooses a signal even though it is not associated with the target referent in their language (0.06), ε is chosen to ensure that the probabilities sum up to 1.

Note that in this phase agents do not distinguish the hypotheses in terms of context-type, but only in terms of the language. In addition, agents are not aware that these data points might have been generated by a pragmatic speaker, so they just consider

that any of the words that could convey the target meaning in the language are equally likely to have been used.

For the first generation in the chain, the pairs correspond to an initial language, which in this case is degenerate (i.e. {{0, 2} : aa, {0, 3} : aa, {1, 2} : aa, {1, 3} : aa)}), so the agents would have to introduce structure starting from scratch. As previously described, subsequent generations learn from the pairs produced by the previous generation. If the size of the data set is larger than the number of learning rounds, the training pairs used in this phase will be selected at random, otherwise the agents are prompted to learn at least once from each pair in the set. Naturally, more learning rounds lead to a more consistent transmission of languages between generations. To ensure stability in the transmission process, we have set this parameter as ten times the number of pairs in a language: $n_{learning} = 40$.

3.5.3 Communication phase (speaker)

The communication phase also takes place over a number of rounds: the higher this number, the higher the influence of the communicative pressure. We experimentally determined that a number of $n_{communication} = 60$ achieves a good balance between pressures. For each round, two agents are randomly selected to play the roles of speaker and listener. First, the speaker sp samples a hypotheses h from its distribution A_{sp}. This will contain the language that the agent will use for communication l, as well as the context-type that it believes the listener is situated in (given previous experience) t. The way these parameters are inferred will become clear at a later point. Next, the pragmatic speaker chooses an utterance u to express the given meaning m to the pragmatic listener, by sampling from a distribution S_1. To determine S_1, the speaker reasons about how a literal listener would interpret each possible utterance in its respective context, then weights its production so as to prioritize those utterances that are the most likely to be interpreted by the listener as conveying the intended meaning.

3.5.3.1 Shared context

When the speaker already knows the context in which the listener will interpret its utterance, it can tailor its utterance to that specific context:

$$S_1(u \mid m, c_l, l) \propto L_0(m \mid u, c_l, l) \tag{3.3}$$

$$L_0(m\,|\,u,c_i,l) = \begin{cases} \frac{1}{|p_u|} - \varepsilon, & (m,u) \in l \quad and \quad m \in c_i; \\ \varepsilon_{err}, & (m,u) \notin l \quad and \quad m \in c_i; \\ 0, & m \notin c_i \end{cases} \tag{3.4}$$

where $|p_u|$ represents the number of meanings that map to the utterance u in the language l and are part of the context c_i, ε_{err} is the probability that the literal listeners makes a mistake and chooses a meaning even though it is not associated with the received signal in their language (0.06), ε is chosen to ensure that the probabilities sum up to 1.

3.5.3.2 Unshared context

If the speaker has no actual way of knowing what specific context will be used by the listener to interpret the utterance in the current round, then it will have to be informative on average with respect to the sampled context-type (i.e. it will compute a separate likelihood for each context of that context-type and then average these out):

$$S_1(u\,|\,m,t,l) \propto L_0(m\,|\,u,t,l) \propto \sum_{c_i \in C(t)} P(c_i|t)\, L_0(m\,|\,u,c_i,l) \tag{3.5}$$

$$P(c_i|t) = \frac{1}{|C(t)|} \tag{3.6}$$

$$L_0(m\,|\,u,c_i,l) = \begin{cases} \frac{1}{|p_u|} - \varepsilon, & (m,u) \in l \quad and \quad m \in c_i; \\ \varepsilon_{err}, & (m,u) \notin l \quad and \quad m \in c_i; \\ 0, & m \notin c_i \end{cases} \tag{3.7}$$

where $C(t)$ is the set of all the contexts that are part of context-type t; all other parameters are defined as in the shared context setup.

Thus, from the perspective of the speaker, the probability of a meaning m being chosen by the literal listener in a context of type t is equal to the product of the probability of meaning m being chosen for a given utterance in a context c_i (using language l) and the probability of that context being the actual context of the listener, summed over all contexts c_i that are included in context-type t.

3.5.4 Communication phase (listener)

After the pragmatic listener ls receives the speaker's utterance u, it samples its own language l from distribution A_{ls}, then has to guess the intended meaning m, by sampling from a distribution L_1. For this, the listener has to reason about how likely

the pragmatic speaker would be to express a certain meaning using the emitted utterance, considering the given context, then prioritize the meanings with the highest likelihoods:

$$L_1(m \mid u, c_l, l) \propto S_1(u \mid m, c_l, l)\, P(m) \tag{3.8}$$

$$m\prime \sim L_1(m \mid u, c_l, l) \tag{3.9}$$

where $P(m)$ is a flat prior over the meaning space, as the listener would not have a bias for any specific meaning

Finally, the pragmatic listener receives feedback and updates its posterior over the hypothesis space, effectively trying to infer the language that was used by the speaker to produce the given utterance. The data used for the update depends on communication success: if the agent identified the correct meaning, it learns from the pair (u,m), otherwise it learns from all the other possible pairings of u and a referent n that was in the context c_l. That is because the agent is not informed about the actual intended referent, so it will have to suppose that any meaning other than the one it selected could have been the correct one. The feedback given to the listener has one more important function: helping the agents figure out the context-type that the listener role is situated in. This would seem rather impractical, since the listener already knows the context that it was presented with. However, even though this inference is done by the listener, no information that is specific to the listener role is actually used for it. Consequently, this situation is functionally equivalent to one where the speaker agent would be doing this inference (i.e. as in the experiment): the agent reasons about how likely the (u,m) pair would have been to result in successful communication for each of the contexts that form a particular context-type.

Thus, after each communication round $k \geq 1$, the agent will update its distribution from round $k-1$ (the distribution after the learning phase if $k = 1$) by considering how likely each of the utterance-meaning pairs (u,n) (as defined above) would have been to lead to successful communication given language l and context-type t. From their perspective, the probability of an utterance u being chosen by the pragmatic speaker to refer to a meaning n using language l is equal to the product of the probability of utterance u being chosen for that meaning in a context c_i and the probability of that context being the actual context of the listener, summed over all contexts c_i that are included in context-type t.

$$
\begin{aligned}
A^k_{ls}(h=(l,t)\,|\,D,u) &= \prod_{n\in D} A^k_{ls}(h=(l,t)\,|\,n,u) \\
&\propto A^{k-1}_{ls}(h)\prod_{n\in D} S_1(u\,|\,n,t,l) \\
&\propto A^{k-1}_{ls}(h)\prod_{n\in D}\left(\sum_{c_i\in C(t)}\frac{1}{|C(t)|}S_1(u\,|\,n,c_i,l)\right)
\end{aligned}
\tag{3.10}
$$

$$
D=\begin{cases} c_l-\{m\prime\}, & m\prime\neq m \\ \{m\prime\}, & m\prime=m \end{cases}
\tag{3.11}
$$

The agent will use this information when it eventually gets the speaker role and has to sample from the hypothesis space, which also means deciding on the context-type that it will be designing its utterance for in the unshared context case.

3.6 Results and discussions

We analyze the simulation results in terms of three metrics: posterior distribution of languages, posterior distribution of inferred context-types and communicative success rate. The first two metrics measure the proportion of the agents' posterior probability occupied by the languages of each separate type (i.e. degenerate, one-feature-only, compositional, hybrid, holistic), and by each of the two inferable context-types (i.e. one-feature-different, mixed), respectively. The success rate tells us the percentage of rounds from the communication phase in which the listener agent identified the intended referent.

3.6.1 Communicating with shared access to context

First, we simulate the more straight-forward situation, where speakers have full access to the communicative context. Since the speaker knows the exact context in which its utterance will be interpreted by the listener, it can also determine, on a round-by-round basis, the distinction that it has to make with its utterance in order to successfully convey the intended meaning.

If the referents that are part of the communicative context can be differentiated using the same feature across all rounds, then senders can use their awareness of the context in order to easily figure out the most efficient communication strategy: conveying

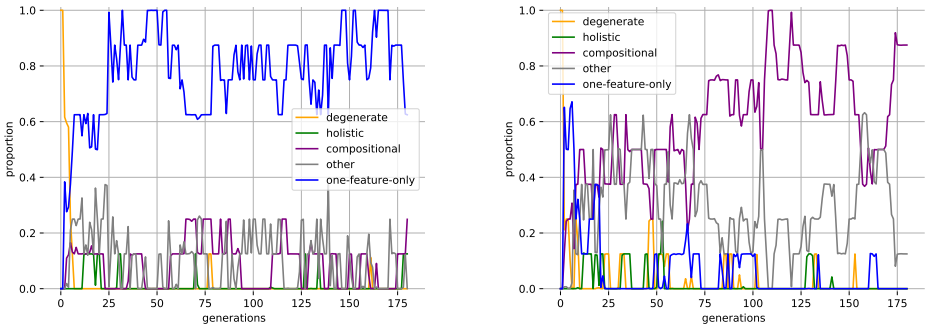

Figure 3.3: Proportion of the agents' posterior probability occupied by the languages of each type, by generation, averaged across 10 different chains, in a shared one-feature-different setting (left) and a shared mixed setting (right).

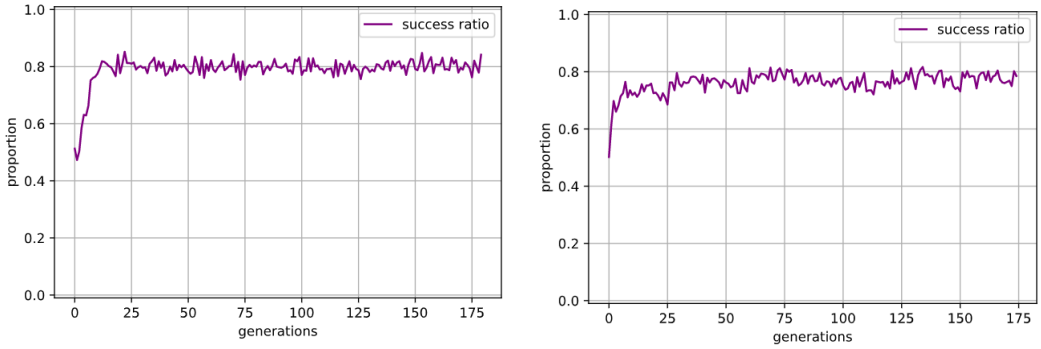

Figure 3.4: Communicative success rate, per generation, averaged across 10 different chains, in a shared one-feature-different setting (left) and a shared mixed setting (right).

only that particular feature in their linguistic system, while leaving out the other. Analyzing the results of the simulation, we first notice from figure 3.4.(left) that the agents realize within the first few generations that the degenerate language they acquired in the learning phase is unsuitable for communication, and gradually move towards more successful strategies. This does not happen instantly in the first generation, but gradually, since the prior of the initial degenerate language is very high, and the agents of a single generation are not communicating for enough rounds to completely push the language towards structure. However, while the transition towards strategies that provide optimal communicative success happens quite early in the chain, it takes significantly more for the agents to stabilize on the optimal strategy, which in this case corresponds to the one-feature-only languages. This is why we see in figure 3.3.(left) a combination of one-feature-only and hybrid languages dominating concurrently, before the more optimal type of language eventually overtakes the other.

In the mixed context-type setting, using the previous strategy no longer guarantees communicative success, as the distinctive feature that has to be communicated by the speaker is no longer consistent across rounds. Indeed, we see in the simulation results in figure 3.3.(right) that, while having a significant proportion of the posterior distribution for the first couple of generations, one-feature-only languages eventually become surpassed by compositional and hybrid languages. The initial rise of the one-feature-only languages can be attributed to them having a significantly higher prior, while also guaranteeing success in communication for over half of the possible contexts. However, as agents observe more data, languages offering more expressive power start to take over. While compositional languages hold the largest share of the distribution at the end of the chains, hybrid languages also prove to be surprisingly competitive throughout the simulation. We suspect that the size of the language space also plays an important role in this result, as the hybrid languages that offer perfect communicative success still share a significant proportion of the prior. However, as the size of the contexts and of the languages themselves would grow, we would expect the systematicity of those particular hybrid languages to significantly decreases, along with their learnability by humans. According to this intuition, compositional languages offer a better balance between simplicity and communicative power as the size of the language space increases. This would however need to be tested. In terms of communicative success, we can infer from figure 3.4.(right) that, again, later generations have very high rates, while performing only slightly worse overall compared with the one-feature-different setting.

It is worth noting that the optimal success ratio is only around 80% for all our simulations, as a consequence of the probabilistic behaviour of our agents. To illustrate, let's assume that both agents are using a fully expressive language, $\{\{0, 2\}$: aa, $\{0, 3\}$: ab, $\{1, 2\}$: ba, $\{1, 3\}$: bb$\}$), and that the speaker is trying to figure out what utterance to chose in order to communicate meaning $\{0, 2\}$ to the listener, within context $c_l = \{\{0,2\},\{0,3\}\}$. The speaker will reason about the behaviour of a literal listener, and will figure out that the signal *aa* will almost always be correctly interpreted, while *ab* will almost always be incorrectly interpreted. However, the other two possible signals are associated with referents that are not part of context c_l, so the literal listener would be equally likely to interpret them as conveying the correct or incorrect meaning. After the pragmatic speaker weights its options, it will most often chose to convey *aa*, very rarely *ab*, and occasionally *ba* or *bb*. Since the pragmatic speaker has no way of differentiating between the two latter signals, it will sometimes make the wrong

choice when presented with a signal other than *aa*, thus explaining the ceiling success rate (which in our case can be mathematically proven to be centred around 0.814).

3.6.2 Communicating without shared access to context

When the speaker holds no information about the distractors in the listener's context, overall predictability greatly decreases. Agents can no longer be sure of the context in which their utterance will be interpreted, so they naturally have to turn towards producing more autonomous and less context-dependent utterances.

While a one-feature-different context-type setting still means that the most efficient strategy for the speaker is to only communicate the distinctive feature, it is no longer straightforward to figure this strategy out. Agents must now try to infer the context-type of the listener over the span of the communication phase, which introduces a level of uncertainty. At the same time, the agents have no means of figuring out the exact context of communication. This introduces a second level of uncertainty, since the speaker's only option is to design a more general utterance, which attempts to help the listener identify the intended meaning in any of the contexts forming the context-type that the speaker inferred. Consequently, we see in figure 3.6.(left) that the languages that emerge are more expressive overall compared to the shared one-feature-different situation: a higher proportion of compositional languages, and a lower proportion of one-feature-only languages. As observed in figure 3.5.(left), this happens because the agents cannot consistently infer the correct communicative setting, so in some cases speakers are designing their utterance for a mixed context-type intead. As a result, the expressiveness of the emerging system is somewhere in between the systems that evolve when access to the context is shared. The presence of both levels of uncertainty turns out to be essential for this to happen: we found that if the speaker has access only to the listener's context-type, there is no need for it to know the specific communicative context for less expressive systems to emerge once again.

Similar dynamics are also at play in a mixed context-type situation: while compositional languages generally hold the largest share of the posterior throughout the chain, the added level of uncertainty determines the speaker to sometimes infer the incorrect context-type (figure 3.5., right), which means that one-feature-only languages also evolve to a lesser extent (figure 3.6., right). As a result, the overall systematicity of the emerging systems is lower compared to the shared mixed context-type case. Since some of the possible contexts that the listener can see here are also present in the one-

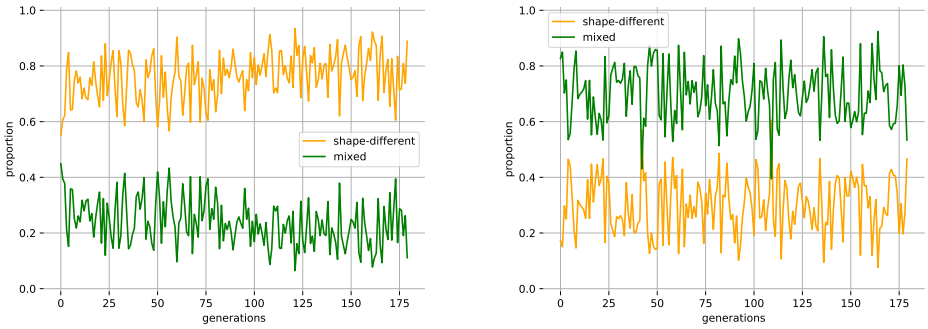

Figure 3.5: Proportion of the agents' posterior probability occupied by each context-type, by generation, averaged across 10 different chains, in an unshared one-feature-different setting (left) and an unshared mixed setting (right), indicating the agent's belief about the context-type in which the listener role has been placed.

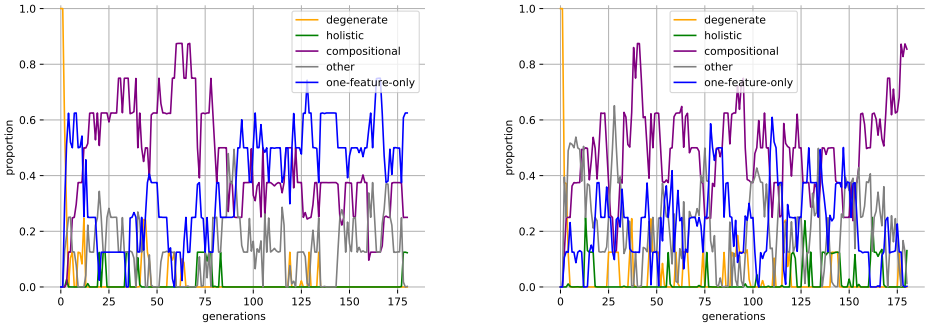

Figure 3.6: Proportion of the agents' posterior probability occupied by the languages of each type, by generation, averaged across 10 different chains, in an unshared one-feature-different setting (left) and an unshared mixed setting (right).

feature-different context-type, there is also a significant chance that the listener will get the correct meaning of the speaker's utterance, even though the speaker tailored that utterance for the incorrect context-types. For this reason, we sense that testing the model for larger contexts might result in higher overall rates of systematicity.

Communicative success for both context-type settings (figure 3.7) throughout the chain is consistently lower compared with the associated shared access to context situations. This is expected, given that the task is now significantly more difficult for the speaker, as inference of the context-type is also necessary. This causes some of the later generations of our chain to perform worse than others, even though they might have inherited perfectly functional languages from the previous generation. Because of this, we also see greater fluctuations in success rates across the chain, for both conditions.

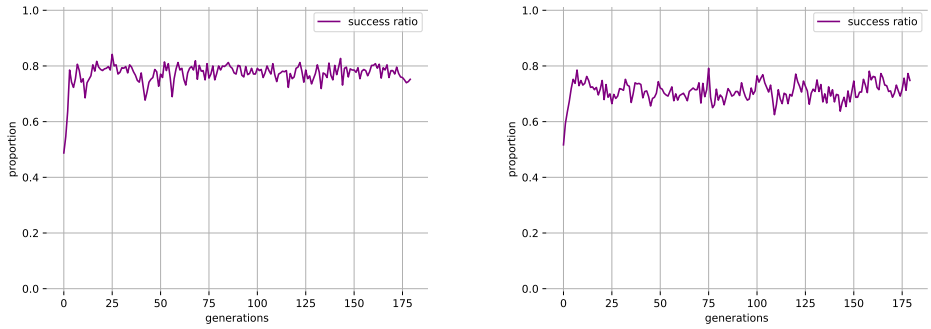

Figure 3.7: Communicative success rate, per generation, averaged across 10 different chains, in an unshared one-feature-different setting (left) and an unshared mixed setting (right).

3.7 Comparison with experimental findings

We now direct our attention to one of the initial objectives of our work: modelling the artificial language experiment in Winters et al. (2018). Before drawing a comparison between the two works, we should note that there is one significant difference in the methodology: the laboratory experiment is not iterated, meaning that its results are achieved by a single generation of participants. The agents of our model are more conformist in their actions, so a single generation of agents does not shift the communication system away from its original form as much as the human participants do. Aside from being a consequence of the RSA framework itself, this difference in behaviour could also be attributed to differences in the design of the learning phase: in comparison with the agents in our model, participants in the Winters et al. experiment are not shown the signals associated with every possible referent, so their data is ambiguous between multiple languages and even types of languages (e.g., they only see the pairs $\{0,2\}$: aa and $\{1,3\}$: bb), resulting in more uncertainty over the initial language. In our case, subsequent generations are needed to slowly push the system towards more efficient and accurate communication. As such, our version is more appropriate for modelling cultural transmission, and can explore the role of contextual predictability on a larger, rather than an immediate, time scale.

In terms of communicative success, we confirm the result that contextual predictability is a good predictor of communicative success: trials were the speaker has access to the listener's perspective yield higher success rates than trials were that is not the case; trials in the one-feature-different condition yield higher success rates than tri-

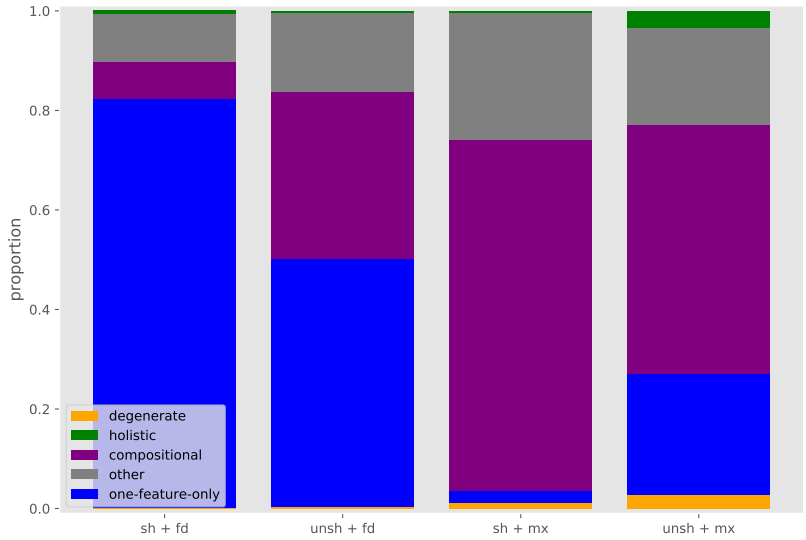

Figure 3.8: Comparison between the total proportion of the agents' posterior probability occupied by the languages of each type, in the last 100 generations of the simulations (i.e. generations 80 - 180), averaged across 10 different chains, for all 4 conditions. In order from left to right these are: shared one-feature-different, unshared one-feature-different, shared mixed, unshared mixed.

als in the mixed condition. Furthermore, communication systems are generally stable in terms of communicative success: later generations do not switch back to strategies that would result in lower success rates.

Winters et al. showed that contextual predictability is also a good predictor for the degree of signal autonomy that will emerge in a communication system. By iterating this process, we can see that, on an evolutionary timescale, contextual predictability affects the systematicity of the evolved languages. However, we find a somewhat less straight-forward relation between the two factors (see figure 3.8): a shared one-feature-different context condition yields a categorical domination of the one-feature-only languages, and almost no compositional languages after a while. Systematicity increases significantly in the unshared one-feature-different context condition, and even further in the two mixed context conditions. However, we find that the shared mixed context condition yields a higher proportion of compositional languages (and as a result, of systematicity) than its unshared counterpart. This appears to be a result of our strategy of modelling speaker behaviour, as well as of the dynamics of iterated learning. We

recall than when access to the communication context is not shared, the speaker's strategy is determined by its belief of the listener's context-type, and each fresh generation of agents starts with a neutral set of beliefs. In the case of just two agents interacting (as in the experiment), the speaker only strengthens its belief throughout the condition, so successful communication strategies will be quite stable. However, when more generations of agents are involved, it is very likely that there will be some generations where speakers infer the context-type incorrectly more frequently, and, as a result, communicate using less autonomous signal. This perturbation causes the language that gets transmitted to the next generation to be less systematic, making strategies not as stable across chains.

3.8 Extending the model to more complex populations

Even though our previous model supports larger populations, it is only applicable when there is a single listener role in that population. In such a situation, the speaker can come up with a single most-efficient communication strategy, one that meets the needs of all of its possible interlocutors. However, a communication system could face pressure to be adaptable to multiple types of settings, involving individuals with potentially differing needs. This section will look at how our model could be extended to such a situation.

3.8.1 Setup

Previously, our agents could only receive two roles: speaker and listener. However, we can add multiple listener roles that the speaker would have to simultaneously adapt to, roles that could be associated with different context-types. To set this up, we opted to keep the pairwise nature of the interaction in the communication phase, and for each round to select the specific listener role that the listening agent will be placed in.

3.8.2 Hypothesis space

The agents will now have to keep track of the multiple existing listener roles, and form a separate hypothesis for each of them. Whereas a hypothesis previously had the form $h = (l,t)$, containing a language and a context-type, a hypothesis for a setup with n different listener roles will contain n different context-types: $h = (l,t_1,t_2,\ldots,t_n)$. This

means that each time we add a new listener role, the size of the hypothesis space will be doubled.

3.8.3 Model

While the learning phase will remain the same as in the base model, some modifications have to be done to the communication phase.

On the one hand, the speaker will be aware of the listener role held by each of its interlocutors, and will have to use the parameter associated with that role in order to produce a separate signal for each listening agent, specifically tailored for its role. In the shared context case, that refers to the associated context, while in the unshared context case, to the associated context-type of the sampled hypothesis. As suggested by the fact that a hypothesis contains a single language, the speaker will produce all utterances for the interlocutors using the same language.

On the other hand, the way in which listeners update their posteriors after receiving feedback must now depend on the particular listener role they have been assigned: the agent is aware of that role and must infer the context-type associated with it. In that respect, an agent a playing the listener role j will update its posterior A according to:

$$A^k_{[j]}(h = (l, t_1, t_2, \ldots, t_n) \,|\, D, u) \propto A^{k-1}(h) \prod_{n \in D} \left(\sum_{c_i \in C(t_j)} \frac{1}{|C(t_j)|} S_1(u \,|\, n, c_i, l) \right) \quad (3.12)$$

where t_j is the hypothesized context-type associated with listener role j; the rest of the parameters are defined as in equations 3.10 and 3.11.

This way, agent a still has a single distribution A^k_a after k communication rounds, but the way this is updated in each round depends on the listener role of agent a in that round. As such, every time an agent is placed in a different listener role, its associated context-type will exert additional pressure on the agent's inferred distribution over the hypotheses.

3.9 The role of population dynamics

In this section, we go beyond the simple two-person communication game, and use our extended model to explore how the population dynamics in early language-using communities might have determined the linguistic structures that they employed. For exemplification, we imagine a situation in which a member of such a community must

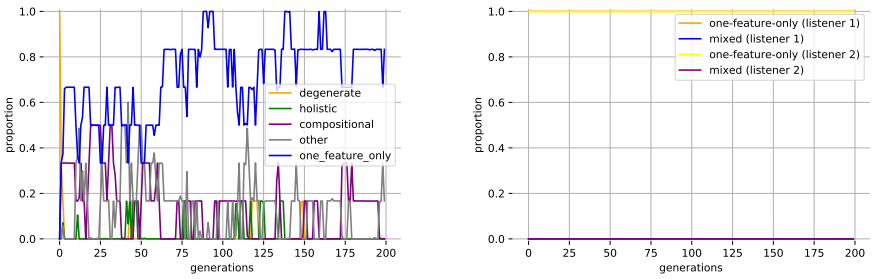

Figure 3.9: Proportion of the agents' posterior probability occupied by the languages of each type (left) and by each combination of context-types (right), by generation, averaged across 10 different chains, in a setup with both listener roles associated with mixed context-types, and agents are highly biased towards inferring one-feature-different context-types (99.99% prior).

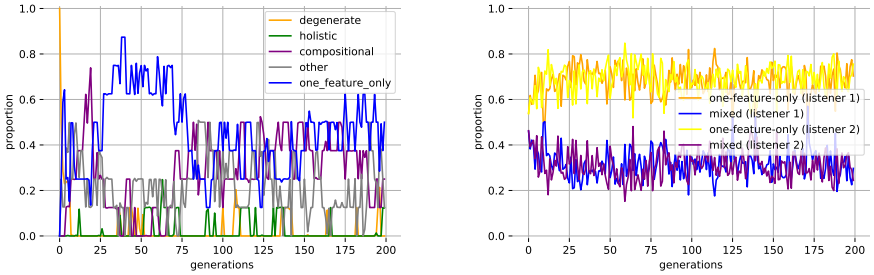

Figure 3.10: Proportion of the agents' posterior probability occupied by the languages of each type (left) and by each combination of context-types (right), by generation, averaged across 10 different chains, in a completely homogeneous group (i.e. one speaker role and two listener roles with one-feature-different context-types associated).

interact with two other individuals, each of them potentially belonging, or not belonging, to that community (i.e. groups of 3).

As outlined in the second chapter, Wray and Grace (2007) propose that the hunter-gatherer communities in which language had first emerged established a linguistic system specifically adapted for esoteric communication. They argue that among the most important factors of this are the homogeneity and small sizes of such communities, which ensured that "most interactions would have been between people who knew each other quite intimately" (Wray and Grace 2007: 568). This tendency towards purely esoteric communication means that not only do all interlocutors share the same cultural and environmental knowledge, but that the also generally take their unified identity for granted.

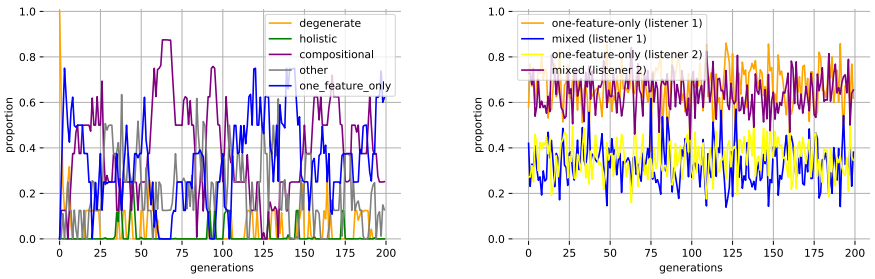

Figure 3.11: Proportion of the agents' posterior probability occupied by the languages of each type (left) and by each combination of context-types (right), by generation, averaged across 10 different chains, in a setup consisting of one speaker role and two listener roles: one associated with a one-feature-different context-types, the other associated with a mixed context-type.

To translate this dynamics to our model, we can imagine that the members of such early communities would have a very skewed prior over the context-type space: they would direct all the prior towards the more specific context-type, which would be the one-feature-different context-type. This is a natural consequence of the assumption that all their possible conversational partners already share some group knowledge (i.e. the one feature that is constant across the communication rounds), which comes from the intimacy of the community. Because of this, the degree of systematicity in their evolved communication system will be low, as shown in figure 3.9., and one-feature-only languages are extremely favoured to emerge. This would appear to have the effect of a barrier to communication with outsiders who do not share the group's knowledge, making it difficult for outsiders to penetrate the society.

As cultural pressure on the community increases (see Kay (1977)), the community becomes more open towards strangers and grows in size, which implies that it naturally becomes harder for its members to know if the individuals that they are interacting with are from their community or not. As a result, the prior gradually shifts towards a more neutral prior, and at some point might potentially become skewed in the other direction.

Our choice of setting a neutral prior over the context-type space would indicate a point in which individuals believe that it is just as likely that a newly encountered interlocutor is a member of their community, as it is for them to be a stranger to the community. After the addition of this extra level of uncertainty, and as individuals start more and more to question their interlocutors' belonging to the community, languages start to exhibit more structure. Consequently, we deduce from figure 3.10. that

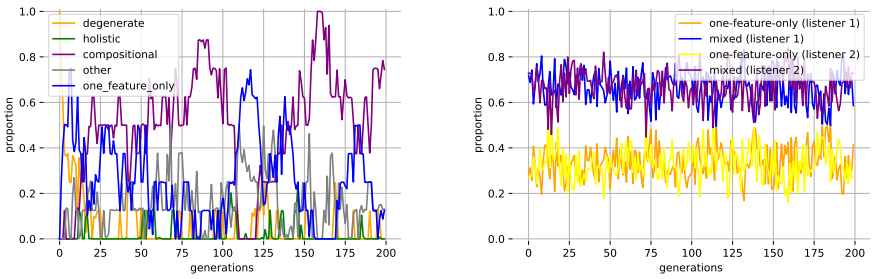

Figure 3.12: Proportion of the agents' posterior probability occupied by the languages of each type (left) and by each combination of context-types (right), by generation, averaged across 10 different chains, in a setup consisting of one speaker role and two listener roles with mixed context-types associated.

even in intra-group communication (i.e. both listeners have a one-feature-different context-type) there is no longer a domination of the one-feature-only languages, but a combination of compositional, hybrid and one-feature-only languages. This would entail that a language losing some of its esoteric qualities is necessary for it to start shifting towards exotericity.

If one of the interlocutors is part of the community, while the other one is not (i.e. one listener has a one-feature-different context-type, while the other has a mixed context-type), we see in figure 3.11. that the communication system evolves to accommodate the needs of both intra-communication and inter-communication. Consequently, overall systematicity slightly grows, but we still see a fair balance between one-feature-only languages and compositional languages.

In our final simulation, both interlocutors are strangers to the community (i.e. both listeners have a mixed context-type associated to them), so the language should adapt even further towards esoteric communication. Indeed, we see in figure 3.12. that the proportion of compositional languages increases significantly compared to the previous settings. We can thus predict that as a community grows in size and becomes more open towards strangers, it will gradually shift its language towards more systematicity on an evolutionary time scale. However, one-feature-only languages also maintain an important, although smaller share of the posterior, indicating that languages cannot completely lose their esoteric qualities as long as the speaker has some uncertainty over the identity of the interlocutors.

We can notice throughout the figures in this section that language-types seem to be quite unstable over cultural time compared to the smaller groups of size two. This

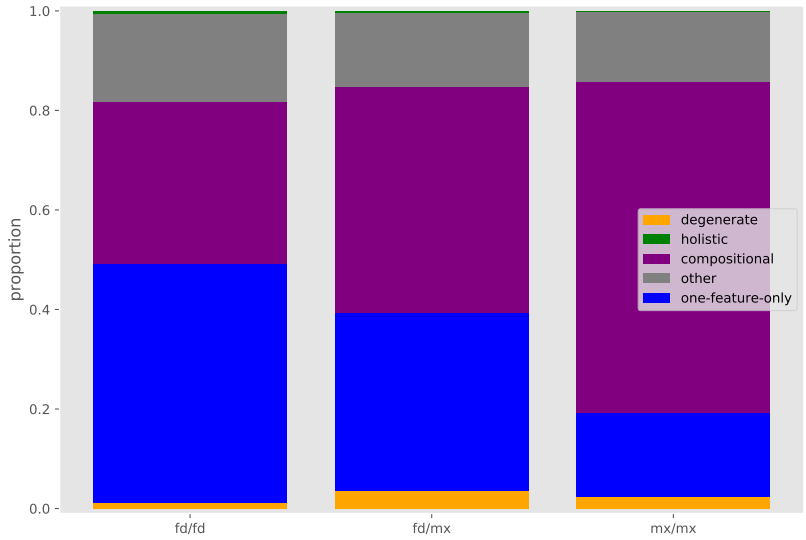

Figure 3.13: Comparison between the total proportion of the agents' posterior probability occupied by the languages of each type, in the last 50 generations of the simulations (i.e. generations 150 - 200), averaged across 10 different chains, for all 3 populations. In order from left to right these are: both listeners with a one-feature-different context-type; one listener with a one-feature different context-type, the other with a mixed one; both listeners with a mixed context-type.

appears to happen because a language now faces possible pressure from multiple interlocutors, so it is naturally more susceptible to inference errors from the side of the speaker. Consider than when there is a single listener, the speaker infers the correct context-types setting in about 70% of the cases. However, as there are now two listeners, for the speaker to infer the correct setting in a communication round, they must pick the correct context-type for both listeners, which only happens around 45% of the time. Notice that this is the case even though the speaker still has a 70% accuracy across the communication phase for each individual listener. As a consequence of this, generations differ even more in terms of the language that they converge on. The mixed context-types would also be in a more advantageous position to be inferred, as they include the contexts that are also present in a one-feature-different context-type. Because of that, if both listeners actually had a one-feature-different context, but the speaker wrongly attributes a mixed context-type to only one of them, a compositional language would emerge to satisfy both types of needs. As such, simulating

using larger populations and manipulating the prior over the context-types for a better balance would be needed to draw more clear conclusions.

3.10 Experimenting with larger languages

In section 3.2 we detailed the structure of the language space that was used in the final version of the model. As mentioned, we used only two different values for each feature that makes up a meaning (the minimal setting that is still capable of demonstrating systematic structure), mainly for the benefits that come with a smaller hypothesis space: lower computational costs. There are however some drawbacks with this approach, as we suggest in some earlier sections: we are constrained to a single possible setting of the one-feature-different context-type (i.e. $\{\{\{0, 2\}, \{0, 3\}\}, \{\{1, 2\}, \{1, 3\}\}\}$). Firstly, this does not allow too much room for exploration for our experiments, since these exact contexts must also be present in a mixed context-type. Secondly, since these contexts contain only two referents, the listener's task is not particularly difficult, as randomly choosing the referent would yield a 50% communicative success rate.

Considering these, we attempted to extend the language space by adding one additional possible value to each of the features, resulting in three values per feature. However, this causes the size of the hypothesis space to drastically increase from 256 hypotheses to over 387 million, as there are now 9 signals and 9 referents (instead of just 4 of each), which can be grouped in a lot of different ways to form a single language. Because of this, it is no longer possible to directly sample from the hypothesis space, so we need an estimation technique.

To obtain random samples $h^\star$ from the probability distribution over our hypothesis space A, we use the Metropolis-Hastings algorithm, which we initialize with a random hypothesis that contains a degenerate language. Following Carr et al. (2020), at every step $i+1$, we select a new hypothesis h_{i+1} by proposing multiple candidates $h\prime$ and accepting one based on its associated acceptance ratio α:

$$\alpha = \frac{A(h\prime \mid D,u)}{A(h_i \mid D,u)} \frac{p(h_i \mid h\prime)}{p(h\prime \mid h_i)} \tag{3.13}$$

Candidates are proposed by sampling from a probability distribution over the hypotheses $g(h\prime \mid h_i)$, which is defined as follows: the language included in the hypothesis h_i is first mutated, by uniformly resampling a number of signal units that is determined by sampling from an exponential distribution (this means that it will be more likely

that a very small number of units will be changed for a mutation); the context-types that were part of h_i are then also resampled, resulting in the candidate hypothesis $h\prime$. This proposal function is symmetric, thus $p(h_i \mid h\prime) = p(h\prime \mid h_i)$, simplifying the formula above.

Next, we generate a random uniform number $u \in [0,1]$ and accept our candidate $h\prime$ if $\alpha \geq u$, otherwise we reject it and keep generating hypotheses until one is accepted as h_{i+1}. To select our effective sample $h^\star$ from the hypothesis space A, we go through a burn-in period and throw away h_1 through h_{2999}. Finally, h_{3000} is selected as our sample $h^\star$.

With this approach, we no longer update the posterior distribution A, which is the main bottleneck of the model, but instead sample a hypothesis for the speaker and the listener when needed using our estimation algorithm. This also means that each agent will only have one associated language and not a complete distribution of languages. As a result, the average amount of time needed for obtaining one sample is reduced from 3 minutes to roughly 30 seconds. Nevertheless, running a single simulation using the parameters mentioned earlier would still take over 24 hours.

Another problem with extending the hypothesis space is that the number of languages of each type increases in a very unbalanced way: the proportion of hybrid languages increases from 89% to over 99.9%. Since re-weighting the priors proved more difficult than expected, partially because tuning the model takes way longer because of the high computational costs, we decided against going forward with the larger language model.

Chapter 4

Conclusions and future work

In this thesis we proposed an extension to the Iterated Bayesian Leaning framework, which replaces the literal Bayesian agents with pragmatic agents based on the Rational Speech Act framework, introduces context-sensitive communication, and provides a model of speaker that is uncertain about the structure of the environment in which communication takes place.

Using our model, we iterate the artificial language learning experiment in Winters et al. in order to investigate how contextual predictability affects the systematicity of languages that evolve over cultural time. We manipulate two aspects of the communicative context: the type of access that the speaker has to that context (shared/unshared), and the historical precedent of the communication act (mixed/one-feature-different). In general, we found our predictions based on results from the laboratory experiment to hold: when the speaker has less contextual information available to exploit, the evolved languages tend to be more systematic. We also found that, in contexts where a less structured language would be sufficient to assure communicative success, speaker uncertainty over the listener's context-type pushes the language towards more structure than is actually needed.

Our model also predicts that languages in groups with homogeneous communicative needs evolve to have less structure than those in groups with more heterogeneous needs. Thus, our results confirm the hypotheses due to Wray and Grace that the communication systems of the earliest language-using communities would have been significantly less systematic than those of contemporary communities. In addition to this, our result show that as a group grows and becomes more outward-facing, it will gradually introduce more systematicity in its language. Nevertheless, some esoteric qualities are bound to be preserved as long as interlocutors have some degree of uncertainty over

the identity of their communicative partners.

We will conclude this thesis by discussing some interesting future directions that could be taken. First, the agents in our model are establishing conventions predominantly by engaging in recursive pragmatic reasoning, which is argued to demand a lot of computational effort from the brain (White et al. 2020). A more cognitively plausible alternative would be to introduce conversational repair mechanisms (e.g., Dingemanse et al. 2015, Macuch Silva and Roberts 2016): one communication round would in this case be made up of multiple turns, so agents could negotiate the meaning of signals in more than a single step by asking for further clarifications when a meaning is ambiguous after the previous turn. This way, the full RSA could be replaced with a less costly, amortized form. Second, one of the important advantages of computationally modelling an experiment is that a model allows us to explore different variations of that experiment. In their work, Winters et al. mention that knowledge of the context is quite limited in their experiment. As such, we could introduce different new types of manipulations to the context, and use the simulation results to guide future possible experiments. In addition to this, their experimental setup forces the speaker to try to figure out the listener's context-type, and only make use of that information when communicating. Similarly, our agents' behaviour is exclusively determined by inferences about their partners' knowledge. However, Lane et al. (2006) argue that interlocutors also consider their own perspective in both production, and comprehension, sometimes even more than their partner's. To accommodate this, we could provide a separate context to the speaker, then change the model of our speaker to consider both the inferred context of the listener, and their own context, in various proportions (e.g., Ryskin et al. 2020), when designing their utterance. We could then see if different proportions result in different rates of systematicity, replicate the setup using a laboratory experiment, and see which proportion better fits the experimental results. Finally, the extensions that we made to the Iterated Bayesian Learning framework (i.e. pragmatic agents, context-sensitive communication, speakers that are uncertain of the structure of the world) also open up exciting prospects for further investigating the role that social factors have on the evolution of language complexity.

References

Barron, A. and Schneider, K. P. (2009). Variational pragmatics: Studying the impact of social factors on language use in interaction. *Intercultural Pragmatics*, 6(4):425–442.

Bergen, L., Levy, R., and Goodman, N. (2016). Pragmatic reasoning through semantic inference. *Semantics and Pragmatics*, 9.

Brennan, S. E. and Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6):1482.

Carr, J. W., Smith, K., Culbertson, J., and Kirby, S. (2020). Simplicity and informativeness in semantic category systems. *Cognition*, 202:104289.

Chomsky, N. (1965). Aspects of the theory of syntax. *Cambridge, MA: MITPress*, (1977):71–132.

Chomsky, N. (1980). Rules and representations. *Behavioral and brain sciences*, 3(1):1–15.

Christiansen, M. H. and Chater, N. (2008). Language as shaped by the brain. *Behavioral and brain sciences*, 31(5):489–509.

Christiansen, M. H. and Chater, N. (2016). *Creating language: Integrating evolution, acquisition, and processing*. MIT Press.

Culbertson, J., Smolensky, P., and Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, 122(3):306–329.

Degen, J., Hawkins, R. D., Graf, C., Kreiss, E., and Goodman, N. D. (2020). When redundancy is useful: A bayesian approach to "overinformative" referring expressions. *Psychological Review*.

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., Gisladottir, R. S., Kendrick, K. H., Levinson, S. C., Manrique, E., et al. (2015). Universal principles in the repair of communication problems. *PloS one*, 10(9):e0136100.

Goodman, N. D. and Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in cognitive sciences*, 20(11):818–829.

Goodman, N. D. and Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in cognitive science*, 5(1):173–184.

Grice, H. P. (1975). Logic and conversation. In *Speech acts*, pages 41–58. Brill.

Griffiths, T. L. and Kalish, M. L. (2007). Language evolution by iterated learning with bayesian agents. *Cognitive science*, 31(3):441–480.

Hawkins, R. X., Frank, M., and Goodman, N. D. (2017). Convention-formation in

iterated reference games. In *CogSci*.

Hawkins, R. X., Stuhlmüller, A., Degen, J., and Goodman, N. D. (2015). Why do you ask? good questions provoke informative answers. In *CogSci*. Citeseer.

Kay, P. (1977). Language evolution and speech style. In *Sociocultural dimensions of language change*, pages 21–33. Elsevier.

Kirby, S. (2001). Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, 5(2):102–110.

Kirby, S., Dowman, M., and Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences*, 104(12):5241–5245.

Kirby, S., Smith, K., and Brighton, H. (2004). From ug to universals: Linguistic adaptation through iterated learning. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"*, 28(3):587–607.

Kirby, S., Tamariz, M., Cornish, H., and Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141:87–102.

Lane, L. W., Groisman, M., and Ferreira, V. S. (2006). Don't talk about pink elephants! speakers' control over leaking private information during language production. *Psychological science*, 17(4):273–277.

Macuch Silva, V. and Roberts, S. G. (2016). Language adapts to signal disruption in interaction. In *11th International Conference on the Evolution of Language (EvoLang XI)*.

Monroe, W., Hawkins, R. X., Goodman, N. D., and Potts, C. (2017). Colors in context: A pragmatic neural model for grounded language understanding. *Transactions of the Association for Computational Linguistics*, 5:325–338.

Müller, T. F., Winters, J., and Morin, O. (2019). The influence of shared visual context on the successful emergence of conventions in a referential communication task. *Cognitive science*, 43(9):e12783.

Nölle, J., Staib, M., Fusaroli, R., and Tylén, K. (2018). The emergence of systematicity: How environmental and communicative factors shape a novel communication system. *Cognition*, 181:93–104.

Perfors, A. and Navarro, D. J. (2014). Language evolution can be shaped by the structure of the world. *Cognitive science*, 38(4):775–793.

Reali, F. and Griffiths, T. L. (2009). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, 111(3):317–

328.

Ryskin, R., Stevenson, S., and Heller, D. (2020). Probabilistic weighting of perspectives in dyadic communication. *CogSci*.

Smith, K. (2006). Cultural evolution of language.

Sperber, D. and Wilson, D. (1986). *Relevance: Communication and cognition*, volume 142. Harvard University Press Cambridge, MA.

Thurston, W. R. et al. (1987). *Processes of change in the languages of north-western New Britain*. Dept. of Linguistics, Research School of Pacific Studies, The Australian ....

White, J., Mu, J., and Goodman, N. D. (2020). Learning to refer informatively by amortizing pragmatic reasoning. *CogSci*.

Winters, J., Kirby, S., and Smith, K. (2018). Contextual predictability shapes signal autonomy. *Cognition*, 176:15–30.

Wray, A. and Grace, G. W. (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, 117(3):543–578.

Yang, C., Crain, S., Berwick, R. C., Chomsky, N., and Bolhuis, J. J. (2017). The growth of language: Universal grammar, experience, and principles of computation. *Neuroscience & Biobehavioral Reviews*, 81:103–119.