



Social Network Limits Language Complexity

Matthew Lou-Magnuson,[†] Luca Onnis[†]

Linguistics and Multilingual Studies, Nanyang Technological University

Received 11 October 2017; received in revised form 25 August 2018; accepted 14 August 2018

Abstract

Natural languages vary widely in the degree to which they make use of nested compositional structure in their grammars. It has long been noted by linguists that the languages historically spoken in small communities develop much deeper levels of compositional embedding than those spoken by larger groups. Recently, this observation has been confirmed by a robust statistical analysis of the World Atlas of Language Structures. In order to examine this connection mechanistically, we propose an agent-based model that accounts for key cultural evolutionary features of language transfer and language change. We identify transitivity as a physical parameter of social networks critical for the evolution of compositional structure and the hierarchical patterning of scale-free distributions as inhibitory.

Keywords: Social network; Language evolution; Language change; Agent-based model; Grammaticalization; Language complexity

1. Introduction

Natural languages vary widely in grammatical structure, especially in the degree to which they make use of nested composition. A graduated divide in this typological space is the extent to which words themselves make use of compositional patterning. That is, the degree on average to which words possess internal structural hierarchies (morphological composition), beyond their arrangement in the phrasal hierarchy of a sentence (syntactic composition). In some languages, such as Chinese, morphological structure is virtually nonexistent, whereas others, like Lushootseed (an indigenous, Salishan language of North America), have so much morphological structure that entire utterances can be a single, highly complex word. It is not yet understood how such typological patterns emerge historically and why natural languages vary in this property over time.

Correspondence should be sent to Matthew Lou-Magnuson, Nanyang Technological University, 50 Nanyang Ave., Singapore 639798, Singapore. E-mail: matt0020@e.ntu.edu.sg

[†]Equally contributing authors.

Computational modeling researchers have proposed both symbolic (Smith, Brighton, & Kirby, 2003) and neural architectures (Batali, 1998) to demonstrate that compositional language structure develops where cultural evolutionary forces are at work (Kirby, Cornish, & Smith, 2008). Indeed, compositional structures appear to evolve as an efficient way to handle the information bottleneck that occurs between generations—a learner must induce the rules of a language from a small sample of observed linguistic usage—emerging as a natural solution to this incomplete information (Tria, Galantucci, & Loreto, 2012). In addition, while non-compositional systems may have been required to initially bootstrap human language, compositionality is preferable once the space of meanings needed to be transmitted grows large; few meanings are easily handled by small iconic expressions, but this strategy becomes untenable over time (Roberts, Lewandowski, & Galantucci, 2015).

This previous work has provided a clear understanding as to why compositionality emerges in communication systems, but languages compose signals in both syntactic and morphological ways. There is no clear preference for one strategy over the other, as evidenced by the even distribution of languages over the spectrum of possible language configurations (Lupyan & Dale, 2010). However, linguists have long noted that rich morphological patterns tend to appear in languages spoken by small groups more than in larger ones (Evans & Levinson, 2009), and some have suggested that smaller social groups are simply better at supporting the kinds of linguistic innovations that lead to these developments (Nettle, 2012; Trudgill, 2011). In addition, languages seem to favor syntactic means over morphological ones as their communities of speakers grow in size (Lupyan & Dale, 2010).

For example, historical records and linguistic reconstruction both posit much more complex morphological patterns for the ancestors of Modern Chinese and English. The set of complex features they each possessed has diminished over time as the speaker population grew and spread. This stands in contrast to their genetically related cousins, for example, Japhug and German, respectively, that have not spread as widely and retained such features (Jacques, 2012; Ringe, 2008). Indeed, empirical evidence suggests the typological patterning that languages display may be connected to aspects of the social network of the speakers.

A recent survey of the World Atlas of Language Structures (Dryer & Haspelmath, 2013) has pushed this observation further to statistical correlation; it was found that after controlling for phylogenetic and areal influence, a novel measure of population spread was highly correlated with the number of grammatical features marked by morphological means (Lupyan & Dale, 2010). Specifically, languages with smaller and more isolated speaker populations tend to make much greater use of morphology than those with larger and more wide-spread populations. Relatedly, evolutionary methods applied to a vocabulary database of Polynesian languages (Bromham, Hua, Fitzpatrick, & Greenhill, 2015) found statistically robust evidence of an influence of population size on the rate of language change, such that larger populations tend to have higher rates of gain of new words whereas smaller populations have higher rates of word loss.

However, beyond correlational evidence, a mechanistic account of how languages gain and maintain complexity is lacking. Prominent sociolinguistic proposals are descriptive in nature (Trudgill, 2011), lacking explanatory power, although computer modeling work from the network theory perspective has made some initial inroads in explaining how demographic factors might affect language evolution (Bentz & Winter, 2013; Dale & Lupyan, 2012; Reali, Chater, & Christiansen, 2018). However, these accounts focus on why languages should become structurally simpler with larger communities and assume an initial level of development by fiat. How morphological complexity was acquired in the first place is not accounted for explicitly by the above studies. A further major limitation with extant approaches is the assumption that the mechanism by which language changes diffuse is epidemic in nature (Ke, Gong, & Wang, 2008). Under these accounts, innovations are like viruses, and contact with someone else infected with the innovation causes one to become infected oneself. This process continues until the innovation spreads completely over the network. However, while language innovation certainly depends on contact with others, language innovation does not spread like a virus but is instead the result of a multifaceted process of social learning. It requires both horizontal diffusion within a population of speakers, and repeated, vertical transmission.

A final limitation of current accounts, and focus of this paper, is that population size and degree of isolation may be conflated and are only a few among several other potential social network variables whose contribution has not yet been investigated (Lupyan & Dale, 2016; Nettle, 2012; Perfors & Navarro, 2014). Similarly, while the processes that generate morphological complexity have been identified in the linguistics literature, where they are studied under the subfield of *grammaticalization*, they have not been explicitly incorporated into existing models. While our model is still unquestionably high level and abstract, it aims to replicate the central grammaticalization mechanism of *reanalysis*.

In this paper, we apply network science to model observations from linguistics and cognitive science, and simulate how human languages may change over time across a social network of speakers. We provide a proof of concept that the topology of a network modulates the developmental direction and depth of linguistic innovation in a given community, everything else being held constant. In particular, we directly test the hypothesis that topologies typical of small human populations promote the development of morphological structures, while those of larger communities lack such capacity, and may in fact lead to inhibitory conditions that encourage the shift to syntactic over morphological patterns. Results from our simulations support this case, thus offering a first causal explanation for the emergence (and not only the simplification) of grammatical patterning as a function of social properties of a community. Importantly, we find that the structural pattern of connectivity within the community is a strong mediator of this interaction. Finally, our results have implications for real-world language communities trying to revitalize or maintain their traditional speech, as the creation of appropriate community relations between speakers may be vital for success.

2. How grammatical complexity develops

Language is transmitted and altered structurally via the mechanism of cultural evolution (Smith et al., 2003). Each learner must observe the linguistic behavior of others and use their endogenous abilities to discover a grammar that accounts for this input (Han, Musolino, & Lidz, 2016). These abilities interject a bias toward certain analyses, and so subtle changes in the input can lead to novel patterns different from those used by the previous generation (Grünwald, 2007; Langley & Stromsten, 2000; Senghas & Coppola, 2001).

In order to model the evolution of structure in language, the assumptions of viral models of innovation must be revised; the gradual, usage-based process through which grammatical changes occur must be accounted for. We aimed to build a model that is sufficiently veridical and thus informed. First, we review the linguistic theory of observed language change, in particular, the role of intergenerational reanalysis as the mechanism that compounds and creates new grammatical structures in natural languages. Then, we suggest a way to model this process, and we estimate the capacity of a network to support such structural development.

2.1. Grammaticalization

The linguistic signals used in human language, both signed and spoken, lie on a gradient that ranges from *lexical* to *grammatical* in kind. Lexical signals characteristically provide labels for objects and their properties, while grammatical signals provide relatively little semantic content and instead coordinate the mechanics of language usage. The linguistics subfield of *grammaticalization* studies the processes by which lexical signals change over time to become grammatical signals. A number of major findings have emerged, including the hypothesis that all grammatical signals originally evolved from earlier lexical origins, as well as the discovery that cross-linguistically there are regular semantic and structural developmental clines (Hopper & Traugott, 2003).

In terms of structural change, signals slowly progress from independent entities to being dependent on other signals for their expression. In general, successive change economizes the physical complexity of the signal and restricts its compositional potential to smaller and smaller subsets of the language. Eventually, this continual reduction in compositional scope and physical expression leads to the signal being unanalyzable by language learners as an independent entity, and eventually it is lost from the language. Important for the current model, such structural changes are almost universally unidirectional (Heine, 2003).

A classic example of the entire grammaticalization process is the history of the future tense from Proto-Indo-European (PIE) down to the Romance languages, here exemplified by Spanish. In PIE, there was no future tense; however, from the PIE desiderative, Classical Latin developed a dedicated future inflection, for example, Latin *amabō* “I will love.” However, this inflexion was lost by the time of Vulgar Latin, and the use of the auxiliary

verb *habere* “to have” became conventionalized to express the future tense; “I will love” had become *amare habeō*, or “I have to love.” In the individual Romance languages, this convention followed language-specific phonetic changes. The progression in the ancestor of Spanish resulted in *amar he** and eventually the auxiliary verb became the future inflexional endings of modern Spanish, *amaré*. Here the process began with an existing PIE inflexional suffix being co-opted for a new purpose in Classical Latin, but it soon reduced in form so much that it was lost. Then a typical lexical item, the verb ‘to have’, is progressively reduced in physical form—*habeō* → *he* → *-é*—while losing semantic content and becoming more restricted in compositionality; it progressed from a free, general-purpose verb to a bound verbal inflexion, simply expressing the future tense. In another several hundred years, it is likely that this inflexion too will be lost, and the process will begin anew.

2.2. Mechanisms of language change

Grammaticalization originally focused on how individual words evolve into grammatical ones, but that is now seen as a specific case of more general processes of language change (Narrog & Heine, 2011). Three interacting mechanisms seem to underly the general process of language change: extension, reduction, and reanalysis.¹ The first two mechanisms are driven by competent speakers of the language, while the latter is driven by language learners.

Extension occurs when speakers use existing patterns in novel ways, often for pragmatic reasons. In the future tense example above, the use of the PIE desiderative to express future time resulted from the novel equating of things we desire to mean things that have yet to happen. The later PIE future with *to have* develops similarly. As mentioned above, such semantic pathways are often attested in other languages, and in fact, English is currently using them both: *I have to love* and *I will love* are both used to express future actions.

Reduction is the second mechanism carried out by competent speakers. As a signal becomes more predictable, it requires fewer distinguishing physical features to be understood. In the example above, as the verb *habere* became more specialized as a sign of the future tense, its occurrence became more predictable from context where future action was implied. With this increase in predictability, less articulatory effort was spent by speakers, and the length of this signal shortened as learners began to internalize the reduced realization.

The final mechanism, reanalysis, occurs when language learners internalize such novel patterns in their input due to extension and reduction by existing speakers. While competent speakers know when a signal is being used in a novel way (extension) or is being expressed more succinctly (reduction), learners do not have such background knowledge of the underlying forms. Instead, there is the chance that these novelties will be taken as fundamental, and the underlying language patterns learned by one generation are different than the one before. In the future tense example, reanalysis occurs when the PIE desiderative shifts from expressing desire for actions that have not yet taken place, to simply

expressing future time without reference to desire. This, again, is precisely what happened with English *will*. Structural reanalysis is also seen above when the reduced form *-é* occurred so predictably after the verb that it became understood to be a suffix rather than a separate element of the utterance.

2.3. *Linguistic reanalysis and structural complexity*

The three mechanisms of extension, reduction, and reanalysis together produce the gradual, directed change from free, lexical signals to dependent, grammatical ones. As signals are extended, learners typically attribute a more general meaning to the signal than was originally understood by prior generations of speakers. For example, the meaning of desire for an action to occur implicitly implies a future time frame for the action. However, as current speakers use it more to imply the temporal meaning than to express their desires, new learners can internalize the temporal meaning as primary, eliding the notion of desire entirely from the use of the signal. In the linguistics literature, this kind of semantic change that makes a signal more general through repeated reanalysis of generalized contexts is called *bleaching* (Hopper & Traugott, 2003), as in some sense the colorful patterns of the signal have been transformed into a generic, white mass.

When a signal is “bleached” in this manner, its frequency of usage increases, as it is able to be applied to meet an increased range of communicative needs. Speakers can more precisely express themselves without implications introduced by generalizing alternative signals that are less bleached, and thus have additional connotations. This increase in usage makes the bleached signal more predictable, as it occurs more and more frequently in similar contexts, and it becomes a target for reduction. As reduced forms of the signal are reanalyzed by subsequent generations as underlying, the signal becomes more likely to be structurally reanalyzed as a dependent element of other signals, usually first syntactically, and then eventually morphologically.

Given the hypothesis that all grammatical signals in language are the result of this grammaticalization process, the amount of reanalysis a signal has undergone provides a proxy for its structural development (Bybee, Perkins, & Pagliuca, 1994; Fortescue, 2016); all bound morphology is the result of repeated cycles of extension and reduction that get cemented through reanalysis. There is no deterministic relationship between the exact number of reanalyses a signal must undergo to become a bound morpheme, but repeated reanalysis is a necessary condition for bound morphemes, and furthermore, the likelihood of being bound increases with the number of repeated reanalyses (Hopper & Traugott, 2003).

3. **Methods**

In this paper, we develop an agent-based model of language change that accounts for the fundamental mechanisms of grammaticalization described above. A language in the model consists of a fixed number of core meanings to be communicated and holistic

signals that exist to express these meanings. Agents communicate by presenting these signals to each other, and depending on their dynamic language experience, either understanding the signal or taking creative action to repair communicability. After a fixed number of exchanges, each agent is replaced by a new one that acquires its linguistic knowledge from the experiences of the agent it is replacing; the new agent acquires the language from the input it receives and forms the next generation of agents.

3.1. Language model

We simulate language as sets of meaning-signal pairings. In all experiments, the agents could express one of three meanings (each represented by an integer value) that stand for a core, grammatical function, for example, tense, aspect, plurality. These meanings are taken to be implicit in the cognition of events, and thus present at all stages of the language, but may vary in the signals used to represent them. Consider the constant need in the history from PIE to Spanish to express the future time of an action but the various signaling means employed over time.

The agents communicate with signals, represented as a sequence of four integer values. The first integer value is the *meaning* that it communicates. The second integer value is the lexical *origin* of the signal. In the PIE example above, two origins would be present: the original lexical origin of the PIE desiderative suffix, and the verb “to have” that is still present in the daughter languages. The third integer value is the number of *reanalyses* through intergenerational transfer that have occurred, and the fourth and final integer value is a unique *identifier* for the signal. As described above, a given reanalysis results in a semantic or structural change to the signal, but a signal representing the same meaning and having the same lexical origin might be reanalyzed the same number of times with different effects. For example, Spanish and Italian share the same general historical developments of the future tense inflexion, but different sound changes took place in each set of speakers (thus there are two languages instead of one) and so even though they share the same lexical origin and similar number of reanalyses, the signals are not identical.

3.2. Agents

An agent consists of four parts: an active repertoire of signals, a passive repertoire of signals, a history of usage experiences, and an identifier. The active and passive repertoires are implemented as sets of signals, where the active repertoire contains signals that the agent uses to communicate with other agents, and the passive repertoire contains all of the active signals, as well as additional signals that the agent comes to understand from others but does not (yet) produce. The identifier is a unique integer value that identifies the agent’s position in the network across generations, and the history is a sequence of pairs, where each pair consists of the identifier of another agent that communicated with the agent, and the signal that other agent used for that communication; the history is a record of who said what to the agent.

Initially, and after each intergenerational transfer, an agent has exactly one signal for each meaning in the language and an empty history. At first, both the active and passive repertoires consist of just these signals and are identical. As communication progresses, described in detail below, each agent comes to understand signals used by others and adds them to the passive repertoire. With additional exposure, signals from the passive repertoire can be added to the active and thus become produced as well as understood.

3.3. Agent-agent communication

An outline of the communication procedure is presented in Fig. 1 (left). Communication proceeds with the “speaker” agent selecting a signal from their active repertoire and presenting it to their partner, the “hearer” agent. Their partner then uses their own passive repertoire to see if they know that signal already, or if the signal is not known, they search the passive repertoire for a significantly close signal of the same meaning they can use to interpret it with. The mechanics of significantly close, explained shortly, capture the fact that while speakers of a language may have different internal understandings of a signal, on the surface their usage is still intelligible to others. Again, consider the future tense example. At one point a generation of Spanish speakers understood the *-é* of the future tense to be the helper verb “to have” while their children or grand children understood that same sound to be simply a verbal suffix that just expressed tense. However, grammaticalization is always gradual, and on the surface all speakers were producing phonetically the same sequences, and always in a context where future time was implied. Something as seemingly different in analysis as a helper verb and suffix can coexist and be mutually intelligible amongst all speakers.

For one of the passive repertoire signals to be sufficiently close, and thus facilitate intelligibility, requires that the integer values of the two signals (the speaker’s chosen signal and the passive one being checked) are comparable. First, the meanings must match.

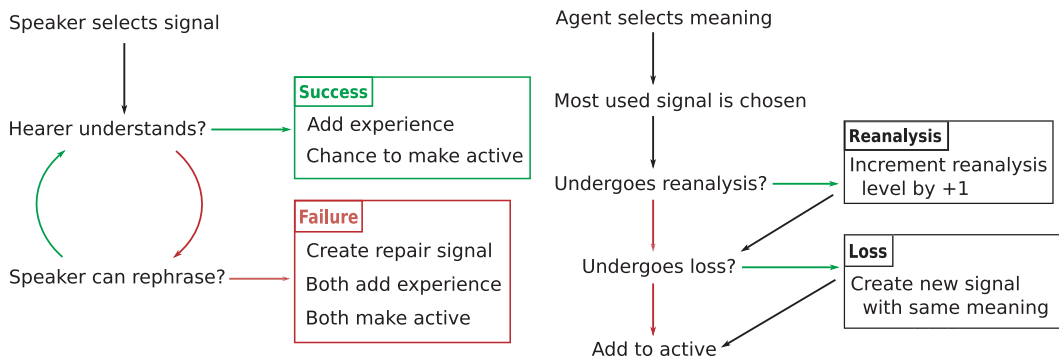


Fig. 1. Communication and intergenerational transfer. The chart on the left, beginning with “speaker selects signal,” is the communication procedure, and the right, beginning with “agent selects meaning,” is the intergenerational transfer procedure. Green arrows and red arrows represent the path taken for yes and no answers to the question in the box, respectively.

Second, the lexical origins must match, so that the fundamental semantic and physical properties of the signal are similar. Finally, the level of reanalyses must be within a set threshold from each other, thus allowing for differences to emerge between the usage of different generations, but not so different that mutual intelligibility is lost. Alternatively, if the signal being used has zero levels of reanalysis, and thus not having undergone any grammaticalization, it is simply a lexical construction with no potential for misunderstanding. Such zero-level signals are immediately understood without need for any specific knowledge from the passive repertoire.

If a signal in an agent's passive repertoire can be found such that the signal being shared with it is intelligible, the agent adds it to the passive repertoire if not already present, and adds a pair consisting of the speaker's identifier and the shared signal to its history. Additionally, if the hearer already possesses the signal in their passive repertoire, that is, it is being exposed again, it undergoes a small chance to add the signal to its active repertoire. In this manner, agents can adapt to and adopt the usage patterns of others.

However, if a signal that facilitates communication cannot be found in the passive repertoire, the speaker will try and repair the communication by using any other active signals it possesses with the same meaning. Again, for each such repair the hearer will attempt to understand it using the available passive knowledge. If no such communicable signal can be found to communicate the intended meaning, the agents will coin a new signal to fill the expressibility gap. Computationally, a new signal with the same meaning as the signal attempted to be shared is created, and given a new origin, a new identifier, and zero levels of reanalysis. This corresponds to the observed behavior in language contact and second language user situations when communication is obstructed by grammatical intricacies. In these situations speakers use periphrastic constructions, that is, whole word phrases, that appeal to adult biases in language understanding (Hickey, 2010; Sebba, 1997). In terms of the model, the new origin captures the lexical content of this new periphrastic construction, and it is given zero levels of reanalysis as it is structurally ungrammaticalized. It should be noted that this is not meant to represent a difference in difficulty on the part of the agents to process signals with reanalysis levels greater than zero. Simply, signals with zero levels of reanalysis have yet to accrue any of the changes that happen during grammaticalization that might prevent them from being understood.

3.4. Intergenerational transfer

After a set number of communication events, each agent undergoes a replacement process that simulates intergenerational transfer, and this is diagrammed in Fig. 1 (right). For each meaning in the language the "child" agent selects a single signal to represent it. To pick the representative signal, the agent searches the history of the "parent" agent it is replacing and attempts to select the signal used by the most interlocutors, that is, the signal that was most generally used to represent the meaning. If there are more than one candidates, then the signal is chosen at random.

After this usage based selection, the mechanisms of extension and reduction are simulated together, as they are linked with respect to the mechanism of reanalysis. Extension

is a constant force in human language, as speakers continually innovate for pragmatic, rhetorical, and descriptive motivations, and as forms become more predictable, their physical structure is elided. These speaker-driven changes directly lead to the possibility that structural changes (whether they are semantic, grammatical, or physical) occur. To account for this, each one of the selected signals undergoes reanalysis as determined by a small base probability for change, representing latent extension and reduction that all speakers interject, multiplied by the number of unique speakers that used the signal in the history. This chance is calculated for each signal that was chosen in the previous step, and each signal probabilistically increases its level of reanalysis based on its individual probability. If a signal does undergo a reanalysis, its reanalysis level is increased by one, and it is assigned the next available identifier to capture the fact it has changed. Again, it should be noted that our model does not attempt to distinguish the nuances of different kinds of structural changes (e.g., semantic bleaching), but rather the more general fact of structural change.

Finally, signals eventually become so reduced in physical form that they become eroded from the language as learners fail to detect their presence. Recall the gradual reduction from the full Vulgar Latin helper verb *habere* to *-é* in Modern Spanish. Eventually, signals are so reduced that they become lost, much like the PIE desiderative suffix that disappeared and created the need to coin the periphrastic construction with *habere* in the first place. To account for this, any signal that is above a set threshold for reanalyses undergoes a chance to be lost from the language. If a signal is lost, the speaker replaces it with a new periphrastic signal, using the same procedure as when two speakers cannot communicate.

Once the new signals have been selected and undergone the mechanisms of language change and loss, the new agent is ready to communicate. It begins with an empty history and with active and passive repertoires that consist of just these chosen signals.

4. Simulation 1: Network transitivity and linguistic reanalysis

The processes of extension and reduction directly feed reanalysis, and repeated reanalysis drives increased morphological complexity. This has led linguists to propose the “intimacy” of small communities as a crucial factor for their high levels of complexity (Nettle, 2012; Trudgill, 2011). While informal, the idea is intuitive. When most speakers know and converse with each other, language learners are exposed to a high degree of variation in the input, which supports rich potential for reanalysis. In turn, such ties allow for quick diffusion of innovations within the community, increasing the likelihood they become adopted and available for repeated reanalysis in the future.

The goal of Simulation 1 is to capture the idea of intimacy mechanistically and in a quantifiable way. We propose that network *transitivity*—also known as the *global clustering coefficient* (Newman, 2010)—is a structural measure from network science that formalizes this concept. Transitivity quantifies the average density of mutual connections within a social network. In a social context, if a network has high transitivity, then the

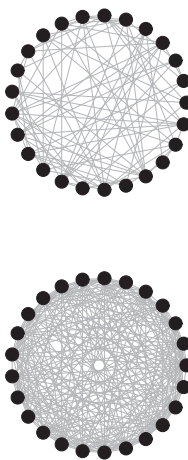
friends of an individual are likely to be friends with each other, as is the case described above in small and “intimate” communities.

Formally, transitivity can be measured in two ways: globally, reducing the entire network to a single density measure, or locally, capturing the density surrounding an individual agent. The global measure is defined as the number of closed triplets divided by the total number of triplets. A triplet is a unique set of three agents and two connections that link them together in the graph. A closed triplet is a unique set of three agents, and three edges that link them together, that is, a triangular relationship among the agents. The local measure of transitivity, defined for an individual agent, is number of edges that exist between the agents neighbors divided by the total number of possible edges that could exist between the neighbors.

4.1. Network architecture and communication

To assess the effect of transitivity, *random networks* (Erdős & Rényi, 1959) were constructed with varying probabilities of connection between agents. Connection probability in random networks is equivalent to average transitivity (Barabási, 2014), and so a spectrum of networks was created with connection probabilities ranging from 1.0 to 0.1, with intermediate step sizes of 10%. Fig. 2 shows two random networks with connection probabilities of 0.7 and 0.2 among a population of 25 agents.

Each network was constructed with 25 agents and used the same initial language across all simulations. Each agent was initialized with the same signal set: one signal for each of three possible meanings, where each signal possessed no initial level of reanalysis; all agents initially used the same, completely periphrastic means of communication. For each connection probability level, 100 unique networks were generated and allowed



Parameter	Value
Reanalysis Threshold	+ - 1
Adoption Probability	$p=0.25$
Erosion Threshold	20
Erosion Probability	$p=0.33$
Variation Probability	$p=0.01$

Fig. 2. Example of random networks used in Simulation 1, with connection probability 0.2 (top) and connection probability 0.7 (bottom). Model parameters are listed to the right.

to undergo 1,000 intergenerational transfer events. Before each intergenerational transfer, agents were allowed to communicate for 10 rounds of communication, where a round of communication consisted of each agent communicating a randomly selected active signal to each of their interlocutors in the graph. The order of these communications was randomized each round to avoid ordering effects.

The communication and intergenerational transfers were conducted as explained in the language model section above, with the specific parameters as given in Fig. 2: Reanalysis threshold is how close the reanalysis levels of two signals, otherwise matching in meaning and origin, could be and still considered intelligible. Adoption probability is the chance at each repeated exposure to a known signal that it would be promoted from passive to active. The erosion threshold is the maximum reanalysis level a signal could reach, above which each time the signal went through intergenerational transfer it was removed with chance equal to the erosion probability. Finally, variation probability is level of latent extension and reduction of the agents; the number of unique speakers experienced using the signal, multiplied by variation probability, yielded the chance a signal would be reanalyzed.

4.1.1. Prediction

As an agent's learning algorithm is usage-based, and sensitive to the kind of signal variation described in the linguistic literature, we expect that network capacity to support sustained reanalysis will correlate positively with increasing transitivity; more intimacy, as operationalized through transitivity, should predict higher levels of reanalysis. This should happen for two reasons. First, high levels of transitivity result in greater variation in each agent's input, and second, they provide denser connectivity that aids dispersal of innovations. Finally, as the statistical correlation of social spread and morphological complexity was linear in the WALS database (Lupyan & Dale, 2010), we would expect a similar linear relationship.

4.2. Results

In order to track the capacity of a network for linguistic complexity, levels of reanalysis were measured in each agent after each intergenerational transfer. At this point in time, directly after intergenerational transfer, every agent has the same number of signals—one for each meaning. The agents have acquired their individual signals from experience, but they have yet to begin communicating. The reanalysis level of every signal used by every agent in a given network was totaled, and then divided by the number of agents, to provide a mean value estimate for the complexity of the language in a particular network at a particular generation. The mean level of reanalysis is displayed in Fig. 3 with each network colored by measure of transitivity. After initial fluctuations in reanalysis values, all networks stabilized by the 500th generation. A mixed-effects model was fitted with individual networks as random factor, and transitivity, as a fixed factor, to predict levels of reanalysis. We selected data from the last 1,000th generation, as we were interested in the stable behavior of the models at the end of their evolutionary path. In line

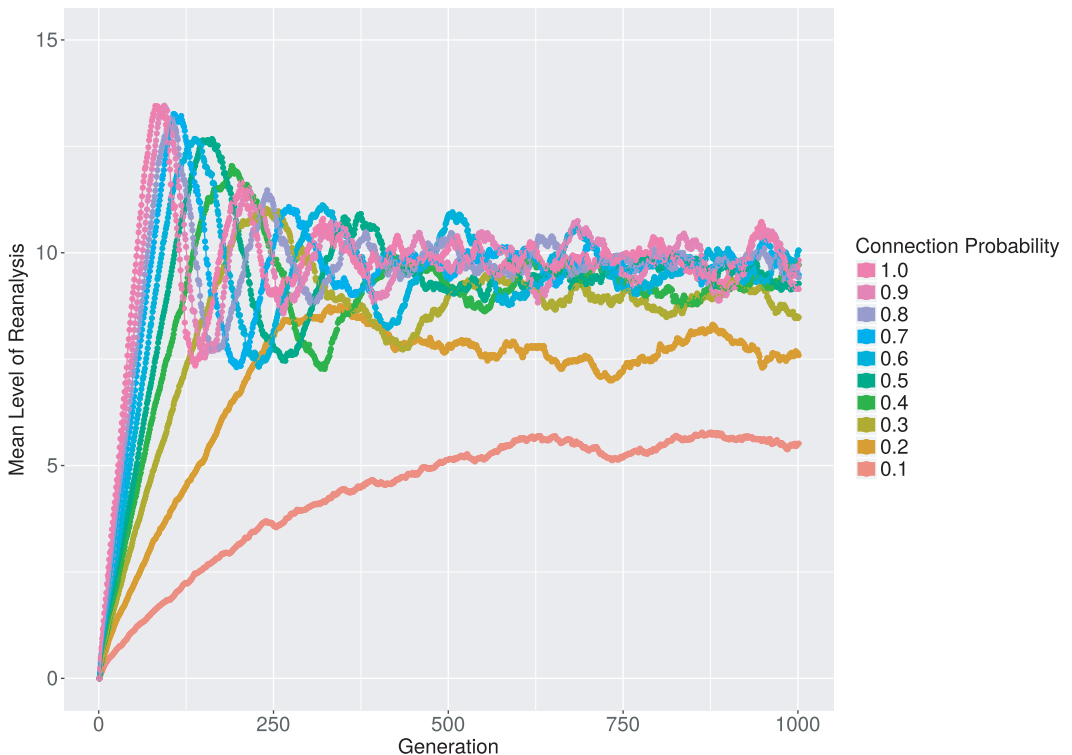


Fig. 3. Results of Simulation 1: Mean levels of linguistic reanalysis as a function of network connectivity over 1,000 generations.

with our prediction, the statistical model yielded a positive effect of transitivity ($\beta = 3.24$; $t = 9.27$; $p < .001$), with more connected networks sustaining higher values of linguistic reanalysis.

4.3. Discussion

As a physical proxy for ‘intimacy’ the effect of transitivity is in line with correlational findings that more intimate societies develop greater usage of bound morphology (Lupyan & Dale, 2010). As network transitivity increases so does the capacity for repeated linguistic reanalysis, and thus the likelihood of morphological complexity. However, the effect of transitivity on linguistic reanalysis is non-linear, evident in the unequal increases in reanalysis despite equal increase in transitivity across conditions. Further, there appears to be a threshold between 0.3 and 0.4 for which networks above the threshold perform qualitatively identical. Above the threshold, reanalysis quickly rises and reaches a high stable state, while in contrast, below the threshold reanalysis stabilizes at far lower levels.

To explore the mechanism driving this finding, we examined the corresponding agent-level measure of connection density, local transitivity. The mean level of reanalysis as a

function of local transitivity is presented in Fig. 4. Each data point is averaged across all networks and agents in Simulation 1 with the given local transitivity measurement.

As seen after one intergenerational transfer (top left) there is virtually no effect of the local density of connections surrounding an agent. However, by generation 100 (top right) the interconnectedness of an agent's immediate neighbors is highly correlated with the level of reanalysis in the signals it produces. By the end of the networks' simulation at generation 1,000, the reanalysis levels have stabilized, as seen in Fig. 3 above. Looking at the local transitivity measures for generation 1,000 (Fig. 4, bottom left), though, we can see that there is still substantial variation between agents, and increasingly so as local transitivity declines. Indeed, averaging over all generations (bottom right) agents with low local transitivity produce signals with lower levels of reanalysis across the lifetime of the network.

It appears that the differences in global levels of reanalysis observed in Fig. 3, and by proxy, overall language complexity, are connected to the local levels plotted in Fig. 4. The distribution of local transitivity by network group is plotted in Fig. 5, with the top row consisting of the two networks immediately above the threshold and the bottom row the two networks immediately below the threshold. In networks below the threshold (bottom row), the majority of agents possess local transitivity values that place them in the

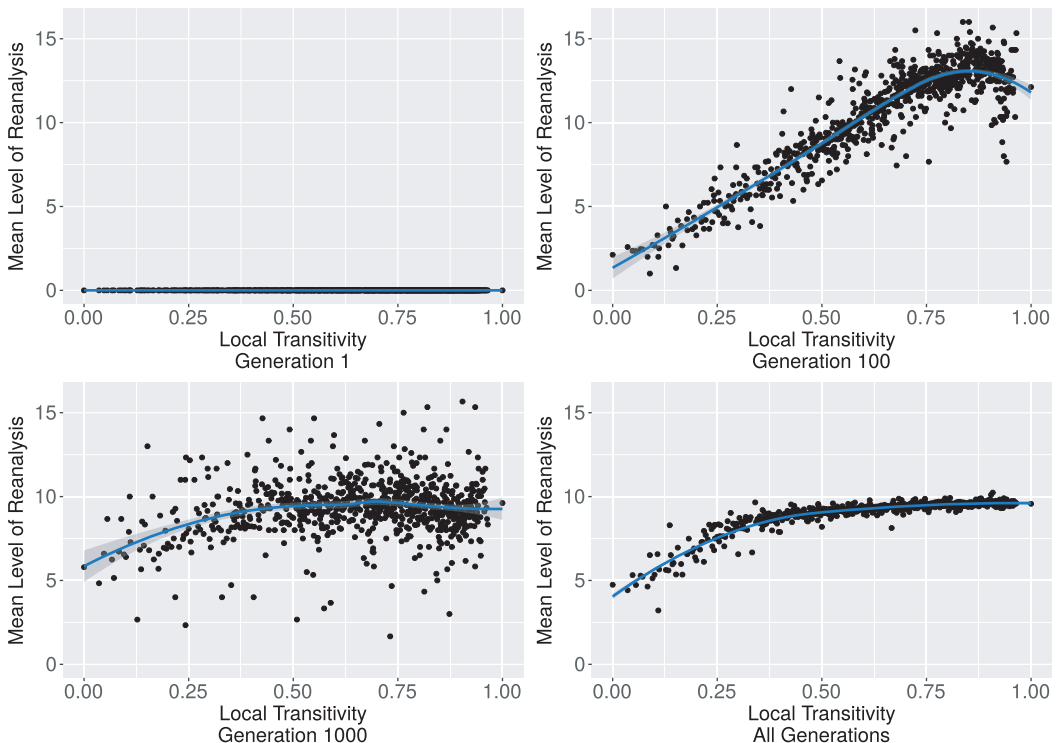


Fig. 4. Local transitivity: Mean levels of linguistic reanalysis as a function of local transitivity. The top two graphs and bottom left are averaged over the individual generation listed, and the bottom right is averaged over all generations.

highly varied region for average reanalysis as seen in Fig. 4. In this region the range of different reanalysis levels among the agents is greater than the threshold for communicability, which causes periphrastic signals to be generated in order to repair this communication gap. Conversely, those networks with global transitivity above the threshold have the majority of their agents with local transitivity values along the stable, constant region for averaged reanalysis.

While the networks with high levels of transitivity are good models of small communities, those with low levels are not. Human social networks are characterized by dense connectivity, especially in small groups, and so the low transitivity networks in Simulation 1 are poor proxies for reality. While these models allow the effect of transitivity to be easily observed in a controlled fashion, the language model must be applied to networks that accurately embody the network properties of both large and small human societies.

5. Simulation 2: Network topology and linguistic reanalysis

As mentioned in Simulation 1, the web of social interaction in small communities is dense, with most members communicating with most other members. As societies grow,

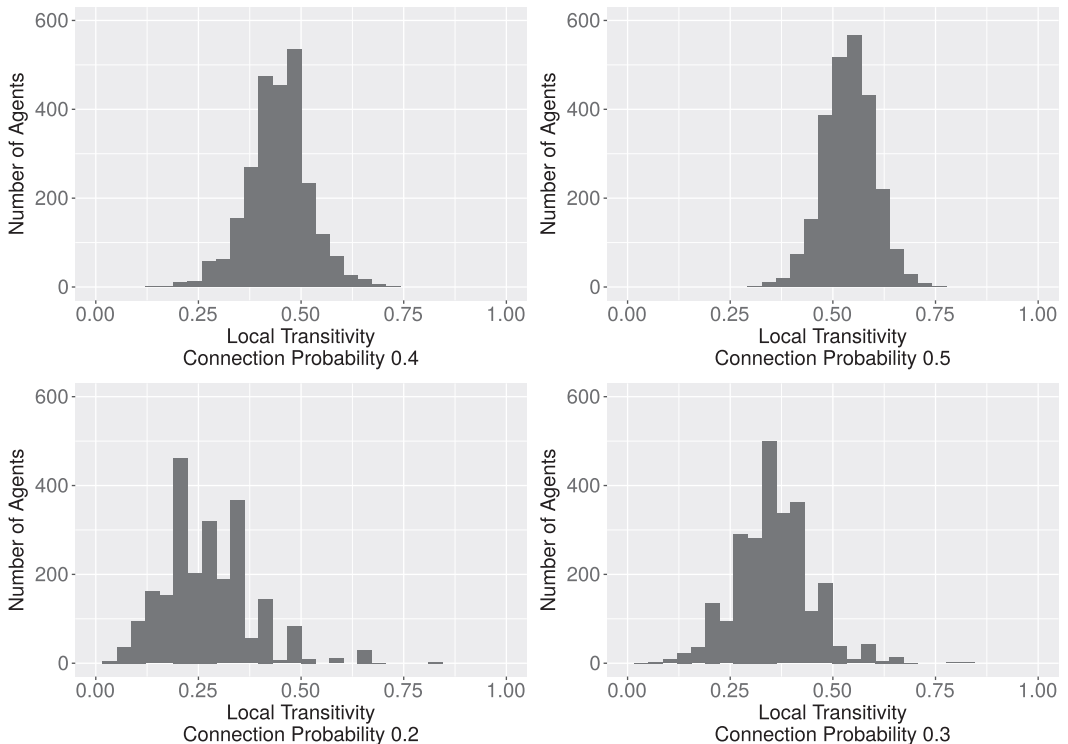


Fig. 5. Distribution of local transitivity: Agent counts are taken from all networks generated at the listed connection probability.

though, they lose this dense connectivity, and communities within communities form (Newman, 2010). People have limits on the amount of time and effort they can input into social interactions (Dunbar, 1998; McCarty, Killworth, Bernard, Johnsen, & Shelley, 2001), and as population size increases, this pressure results in local clusters of social ties (Jin, Girvan, & Newman, 2001). In large linguistic groups, for example, English speakers, people socialize with a small fraction of the whole. Human social networks, both large and small, demonstrate high levels of clustering, but large communities display a major structural characteristic that small communities do not: Large networks are typically scale-free (Newman, 2003; Ravasz & Barabási, 2003).

The scale-free property refers to the fact that on average most agents know only a small percentage of the others, but there exist some agents that know a much larger percentage; there is no set scale in terms of how few or how many connections the typical person may have, and so-called hubs with numerous connections cross-cut locally connected communities. Consider an exemplary scale-free network, that of airports and flights between them. The majority of airports are regional, sending a few flights in a local area, while a few hub airports send the majority of flights across nations and the globe (Barabási, 2014).

5.1. *Network architecture and communication*

We modeled four distinct network topologies. In order to capture the connectivity patterns of small groups, a *complete network* (Complete) was used, where by definition all agents interact with all others; it is analogous to the organization found in a small village. This is the top network of Fig. 6. In contrast, to capture the topology of larger societies, a *hierarchical network* (Hierarchical) designed specifically to mimic the social hierarchy of larger human social networks was used (Ravasz & Barabási, 2003); this is a model of connected communities, analogous to a physically divided network of villages or the socially divided network of a modern city. It is shown in the bottom of Fig. 6. By way of control, two additional networks were tested. The first, *Barabasi-Albert* (BA), is commonly used in viral models of innovation and allows for the creation of a scale-free network with zero transitivity (Barabási & Albert, 1999). By eliminating transitivity from the network, the effect of hub agents can be assessed in isolation. In viral models, the presence of hubs drives the spread of innovation by linking large numbers of agents that otherwise may be distant from each other.

It would be ideal to also compare the common BA model with the transivities present in the hierarchical network; however, a number of problems exist for this at present. First, the traditional BA model does not produce the requisite levels of transitivity present in observed social networks (Ravasz & Barabási, 2003; Varga, 2015). Second, while modified versions of the BA model exist to try and make this property an adjustable parameter (Jin et al., 2001; Varga, 2015), as well as novel procedures to produce scale-free networks with tunable transitivity (Chakrabarti, Heath, & Ramakrishnan, 2017; Herrera & Zufiria, 2011), no method yet exists that accounts for the distribution of connection density in a way that mimics social networks and/or allows for manipulation of that density

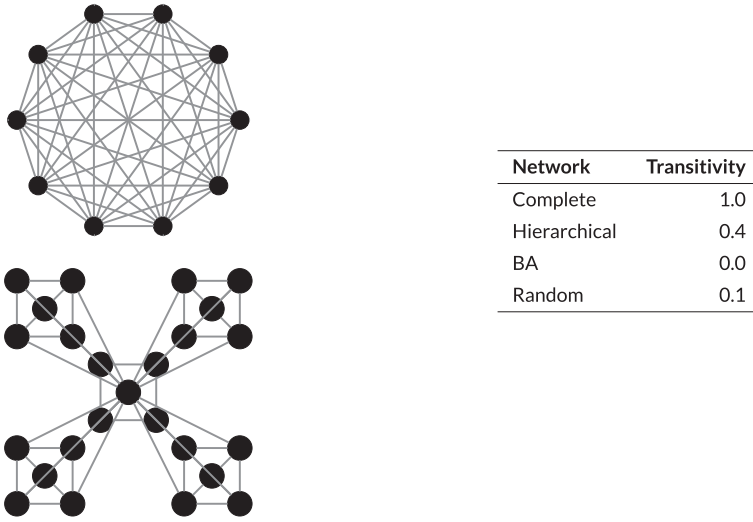


Fig. 6. The complete (top left) and Hierarchical (bottom left) networks used in Simulation 2 exemplify connectivity properties of two common types of human social communities, respectively: the village and the modern city. The transitivity measures for all networks (BA and Random not pictured) are given to the right (the Complete network depicts only 10 agents for clarity).

measure beyond a narrow range. A key issue is that in large, scale-free human social networks, the local transitivity of an agent is inversely proportional to its number of connections, for example, to its degree (Soffer & Vazquez, 2005). At present, generative models of social network structures are an ongoing area of active research, as represented in the above works.

Lastly, a *random network* (Random) was also created (Erdős & Rényi, 1959). The connection probability was $\frac{\ln N}{N}$, the threshold for single connected components, and where N is the number of agents. This allows for the creation of a non-scale-free network with no assumptions about the hierarchical distribution of agent ties.

As with Simulation 1, each network contained a fixed number of 25 agents across all conditions, starting with the same initial language at generation 0. Only the typology was varied. The population size of the agents, the initial language, and the number of communications per agent were the same in all conditions; the only difference was the patterns of connectivity. All simulation parameters were identical to Simulation 1, Fig. 2.

5.1.1. Prediction

The transitivity values for these networks are summarized on the right side of Fig. 6, and both Complete and Hierarchical are above the threshold for maximal linguistic reanalysis seen in Simulation 1, while BA and Random are far below it. If it is the case that transitivity alone affects the ability of a network to support sustained reanalysis, then given these values, we expect Complete and Hierarchical to support similar, maximal levels of reanalysis, and BA and Random to support little to none.

However, Simulation 1 did not account for the kinds of hierarchical structure seen in actual human social networks. If it is the case that these hub connections that develop in scale-free networks are important for supporting sustained reanalysis, then we would expect that BA and Hierarchical to have an advantage over Complete, as both of these networks are scale-free and Complete is not. This is especially so for Hierarchical, which has a transitivity level above the observed threshold for maximal reanalysis.

5.2. Results

Complexity was measured using the same method as Simulation 1, and is plotted in Fig. 7 as a function of generation and network topology. Different paths of evolution of reanalysis are noticeable across network topologies. In particular, Complete exhibits an initial large increase with further stabilization, while Hierarchical and Random proceed along more gradual increases, while Barabasi hardly takes off. To confirm these trends, a mixed-effects model was fitted with individual networks as random factor, and topology at the final generation as a fixed factor, to predict amount of reanalysis. The model yielded main effects of topology, such that the Complete network yielded larger levels of reanalysis. In particular, as predicted the Complete network developed significantly higher reanalysis than the Hierarchical model ($\beta = 4.27$; $t = -315$; $p < .001$). Conversely, the Barabasi model exhibited significantly lower reanalysis than the Hierarchical model ($\beta = -5.09$; $t = -14.82$; $p < .001$). The Random network did not differ from Hierarchical ($\beta = -0.21$; $t = -0.63$; $p = 0.52$).

5.3. Discussion

A seemingly surprising finding is that despite the level of transitivity being above the threshold for maximal reanalysis, the capacity for reanalysis in Hierarchical is no different than Random. A crucial difference, though, is in how that transitivity is distributed within the networks. The high transitivity in Hierarchical is due to the dense connectivity within its five-agent clusters (c.f. Fig. 6), with a transitivity of 0.7 each. However, communication between these small, near-complete subcommunities is mediated through the central agent of the central cluster, the *hub* agent of the network. It is this mediation through the hub that counteracts the complexity generated by these dense clusters.

Zooming in on a single network provides a clearer interpretation. The top of Fig. 8 plots the average level of reanalysis over the 1,000 generations of a single, typical, run of Hierarchical. Characteristic of simulations on Hierarchical is the steady climb of reanalysis levels, as predicted by the high transitivity of the individual clusters, but punctuated by both small and catastrophic declines. While there is always the possibility that a mutually intelligible signal with a lower level of reanalysis may replace another, the small and catastrophic declines are caused when the signals in one group become mutually unintelligible with the others.

It is impossible for the hub agent to learn and actively produce every distinct signal for all five clusters, and thus over many generations the signals within a group

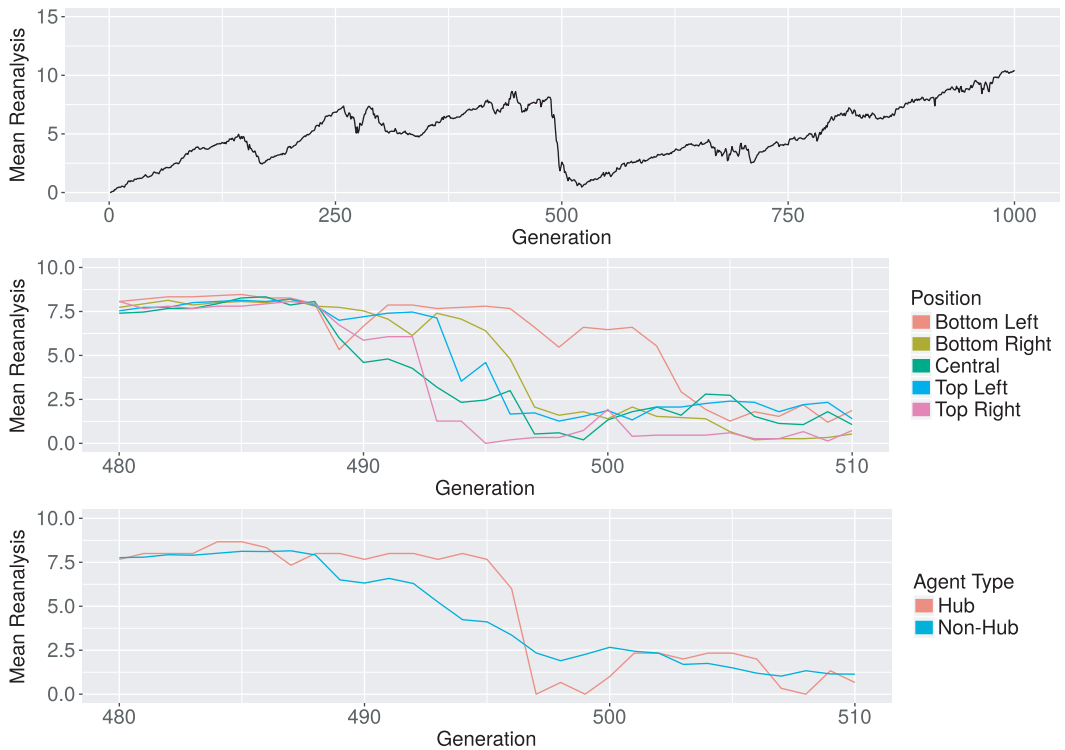


Fig. 7. Study of hierarchical network 3: Top figure displays average reanalysis levels across all 100 generations of the network. The middle and bottom figures center around the large drop in reanalysis at generation 500 and are broken down into the physical clusters of agents (middle) and central hub agent versus others (bottom).

may become locally more complex than those of their neighbors, or become so complex that they erode and are replaced. At this point communication is broken across the groups, and periphrastic signals are generated as the hub tries to maintain communicability. These periphrastic signals, which are automatically understood regardless of passive knowledge, may spread and replace the natural grammaticalization process of another origin. This would be analogous to the *will* future in English becoming dominant and replacing the more reanalyzed *going to* future and its variants, for example, *gonna*, *gun'*, *goin'*, and so on.

Often, this small reduction in average reanalysis is isolated to the central community in which the hub is embedded and one or two local clusters with which communication was repaired, and it only results in a small dip in overall network reanalysis levels. However, the hub is in contact with all five communities, and even when its own central cluster and others are undergoing such a replacement, it may continue to acquire signals with high levels of reanalysis still in use in other groups. This distinct hub behavior leads to catastrophic declines, as occurs around generation 500 in Fig. 8.

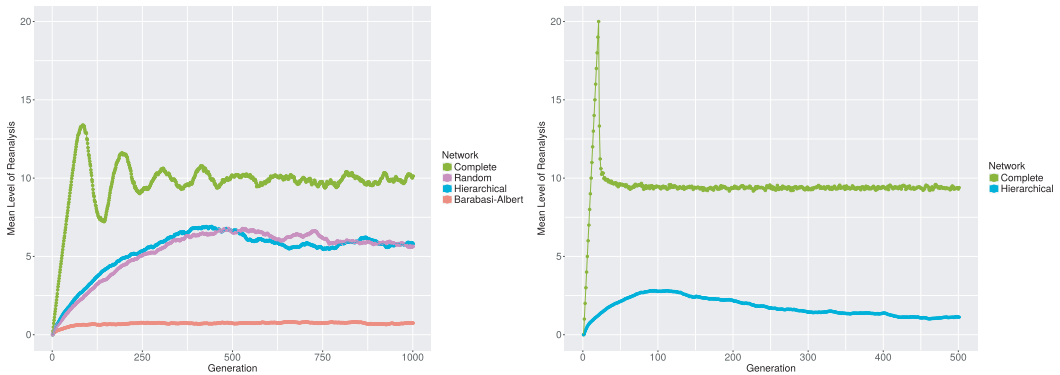


Fig. 8. Results of Simulations 2 and 3: Simulation 2 (left), mean levels of linguistic reanalysis in four different 25-agent network architectures for 1,000 generations. Simulation 3 (right), mean levels of linguistic reanalysis in two 125-agent networks of interest for 500 generations.

The bottom two graphs of Fig. 8 are centered around this catastrophic decline, with the middle graph showing the average complexity in the individual clusters (named for their physical positions in Fig. 6) and the average complexity of the hub agent versus the others. We can see that a repair happens between the central and bottom left groups, causing the sharp decline in average reanalysis as periphrastic signals replace existing ones. However, even as the hub's central group drops ever lower, the hub-agent itself retains high-level signals until all but one group has declined. At this point, the hub agent has caused a number of now competing periphrastic variants to emerge, and rather than quickly recovering, the entire network must undergo a period of little to no growth; compare the quick v-shaped drop and rise after small declines, such as at generations 170 and 210, with the protracted and u-shaped pattern surrounding the catastrophic decline around generation 500. Not until winning variants emerge can they slowly undergo joint reanalysis in the individual clusters.

This innovative behavior of hubs is also predicted by cultural evolutionary theory, as they form *bottlenecks* that restrict the flow of information (Smith et al., 2003). When the transfer of social knowledge is incomplete, there is a pressure exerted to innovate and fill the same cultural and communicative function. The hub agents do precisely this, responding to the pressure caused by gaps in communicative ability, and innovate new signals to fill them. If this constraint to maintain communicability were removed, the five local clusters in Hierarchical would go on to develop individual languages, and the linguistic community would fracture. The price for maintaining cohesion of language as it spreads across such bottlenecks, be they physical or social, may be paid with morphological complexity.

Finally, the periodicity of these collapses was investigated with additional simulations that examined the interaction of two key model parameters: the number of communication rounds before intergenerational transfer, and the multiplier used to determine when an innovation that leads to reanalysis occurs.

One would expect that increasing the rounds of communication between transmissions would reduce the frequency of collapses like those seen in Fig. 8. The chance that two agents are unable to communicate decreases, as each additional round of communication provides more opportunities for innovative signals to diffuse across the network. If agents are more readily able to maintain communication, then fewer repairs will take place, and thus fewer opportunities to trigger a collapse will be present.

Conversely, as the probability that innovations are introduced increases, one would expect that the number of collapses would also increase. As more innovations occur, there are more signals in competition for expressing a given meaning, and this competition reduces the likelihood that agents are able to communicate. With each failure to communicate, a repair is made that may potentially lead to a collapse.

To explore this interplay, the effects of these two parameters were tested with the following values: rounds of communication was varied across 10, 20, and 30 rounds, while the variation probability was tested at 0.001, 0.005, 0.01, 0.05, and 0.01. As a reminder, the simulations reported in the previous sections had 10 rounds of communication before transmission, and as reported in Fig. 2 (right) a ‘variation probability’ of $p = .01$ was used.

In order to calculate the average time between collapses, local maxima and minima were identified using the discrete analogs of traditional calculus. Specifically, the forward difference between points was used to approximate the first and second derivative information. As seen in Fig. 8 (top), the mean reanalysis per generation is quite noisy, with many small rises and falls, and so to de-noise the data only every tenth generation was considered. On this smoothed data, any time a drop in mean reanalysis of greater than 50% occurred, it was classified as a collapse. For each pair of conditions, and additional 100 hierarchical networks were run, and the results summarized in Fig. 9.

As predicted for variation probability, increasing the likelihood for additional competing innovations has the effect of decreasing the average time between collapses. However, in general, the predicted effect of increased communication rounds is not present. That is, as more rounds are afforded to the agents to communicate, the average time between collapses remains constant. The lowest variation probability ($p = .001$), though, does conform to the expectation that increasing communication rounds also increases stability.

It seems then that the stability of such hierarchical networks is most sensitive to the general amount of innovation introduced by the speakers. Moreover, this constraint on stability seems to persist despite increased diffusion of signals across the network. However, it may be the case that if innovation is so gradual as to not cause competition, increased diffusion does translate to increased stability. In particular, for the $p = .001$ case, at 30 rounds of communication before transfer, the average interval between collapses is approaching the length of the simulation. This means that, behaviorally, as communication rounds increase the collapses that characterize the hierarchical networks (as opposed to the complete networks) may be averted.

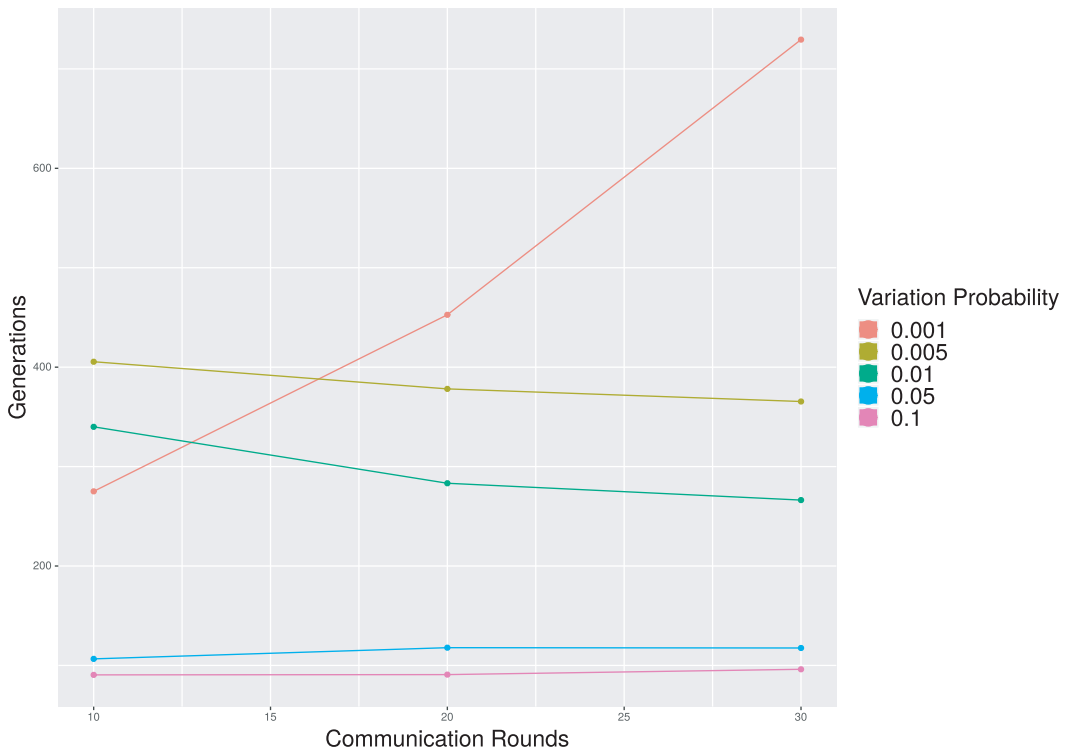


Fig. 9. Interplay of parameters. On the x -axis are the number of rounds of communication allowed before transmission, and the color codes for values of the variation probability, the multiplier used to determine when innovations lead to reanalyses. The y -axis is the mean number of generations between collapses.

6. Simulation 3: Revisiting network size

In Simulations 1 and 2, only the connectivity patterns of the agents were being manipulated—all networks consisted of 25 agents throughout and began with identical languages. While population size has been reported recently as a highly predictive variable in other computational and correlative studies of language complexity (Bromham, Hua, Fitzpatrick, & Greenhill, 2015; Reali et al., 2018), it is also conflated with social structure (Carneiro, 1967; Soffer & Vazquez, 2005), and indeed it may limit the range of possible structures for individual agents (Dunbar, 1998). In particular, the size of complete or near complete networks may be limited to about 100-200 social connections (Gonçalves, Perra, & Vespignani, 2011). In order to examine how our language model may be informed by these findings, we constructed 125-agent versions of the Complete and Hierarchical networks used in Simulation 2. To keep computational time to run on PC computers, we reduced the replications of each network topology from 100 to 50. In addition, because network behaviors in Simulation 2 stabilize after around 300 generations, we also reduce the number of generations from 1,000 to 500. While not discussed here, both Simulation

1 and the additional networks used in Simulation 2 (BA and Random) can be found in the supplementary materials.

6.1. Results and discussion

A mixed-effects model was fitted with individual networks as random factor, topology, and network size (25 vs. 125 agents), and their interactions as fixed factors. The dependent variable was the mean level of reanalysis for each network at generation 500, as seen on the right of Fig. 7. The model yielded main effects of topology, such that the Complete network yielded larger levels of reanalysis ($\beta = 3.65$; $t = 10.79$; $p < .001$). A main effect of netsize also obtained, with 25-agent networks yielding overall larger values of reanalysis over 125-agent networks ($\beta = -5.41$; $t = -12.88$; $p < .001$). Finally, the topology \times netsize interaction was significant ($\beta = -4.51$; $t = 7.71$; $p < .001$). Fig. 10 reveals that driving the interaction was a larger mean difference in reanalysis between Hierarchical and Complete for the 125-agent networks. In addition, mean values of reanalysis for Complete were comparable, if not higher in the 25-agent compared to the 125-agent networks, while mean values of reanalysis were lower for Hierarchical in 125-agent networks. This pattern of results suggests that network size amplifies the effect of network structure not by boosting complexity in larger Complete networks, but by limiting the opportunities for complexification in larger Hierarchical networks.

While the relative performance of these networks at the 125-agent level mirrors their 25-agent counterparts, their trajectory before reaching stability was slightly different. In the 125-agent Complete, signals rose in reanalysis in virtual unison, until becoming so high-level that then began to erode and be replaced. In the 25-agent Complete, the signals did not all rise and erode in such a tight time scale, and so there are some perturbations before the steady state is reached. Looking at the 125-agent Hierarchical, it undergoes a slow rise like its 25-agent counterpart, and peaks as each of its five 25-agent Hierarchical clusters stabilize. However, these clusters themselves participate in a hierarchical arrangement, and the presence of these additional hubs leads to additional declines, driving the overall complexity down again.

7. Conclusion

It has long been posited that dense, “intimate” social ties are responsible for creating conditions that support greater morphological complexity. Insofar as network transitivity captures that idea, the ability of networks to sustain reanalysis of signals over many generations supports this hypothesis. Global transitivity of a network is a reliable predictor for the performance of complete and near complete social networks, which model small and localized populations of speakers. Further, the effect of global transitivity could be seen to reflect different behaviors in the individual neighborhoods of agents, as captured by the measure of local transitivity.

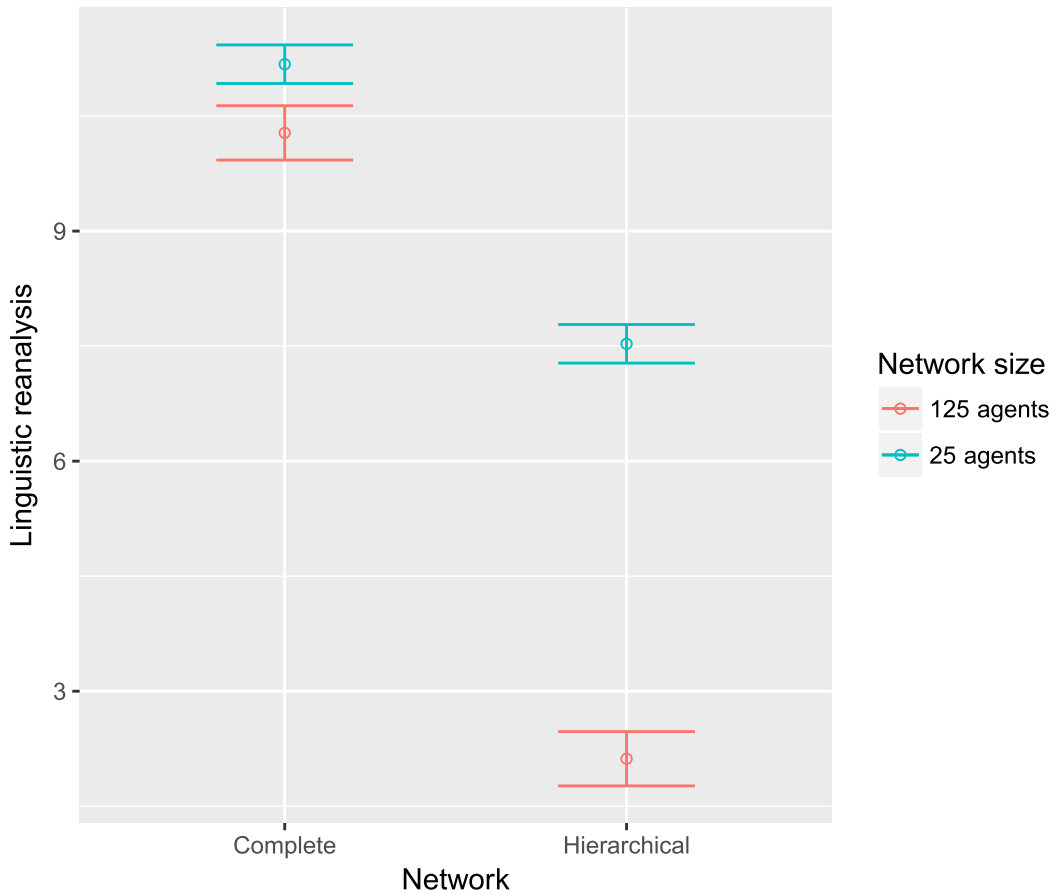


Fig. 10. Interaction between network topology and network size in Simulation 3.

However, as seen in Simulation 2, the capacity for dense connections in a network to support reanalysis at high levels, can be largely overridden as hierarchical structure is introduced. This is contrary to what epidemic diffusion models would predict (Ke et al., 2008) and underscores the need for modeling in network science to closely consider means of propagation, as well as looking beyond questions of “time to spread” or “time to convergence” when considering intergenerational phenomena.

Beyond positing a mechanistic explanation of the observed correlation between morphological complexity of a language and the demography of its speakers, this research may have practical implications as well. Language maintenance and revitalization efforts are increasingly important as languages spoken by smaller languages continue to be lost as globalization prioritizes larger languages of economic and political importance. In particular, many such small languages possess high degrees of complexity which may be best retained if minimum levels of local transitivity are ensured, and hierarchical structure in the domains where revitalization efforts take place are minimized.

The present work represents an initial exploration at the intersection of network science and computational linguistic modeling. For instance, our model generalizes over the distinctions between semantic and grammatical structure, instead focusing on the general process that unifies them. It may very well be the case that further more complex interactions emerge once this level of resolution is added in future work. Also, as clear from the contrast of the hub versus other low local-transitivity nodes in the random networks, the effects of network structure surrounding such measures is also clearly an area requiring further effort.

Further afield, while our simulations focused on the intimacy of speakers as a causal variable, it is entirely possible that structural aspects of natural languages be affected by a combination of other social variables, some of which are discussed at length by Trudgill (2011) and Nettle (2012): community size and degree of isolation; degree of language contact; type of language contact; degree of social stability; density of social networks (modeled here); amount of communally shared information; vocabulary size of the community; ratio of child to adult language learners in the community.

Our findings match the observations made by linguists, as well as the statistical findings derived from the World Atlas of Language Structures on population spread and compositional structure (Dryer & Haspelmath, 2013; Lupyán & Dale, 2010). Small populations with dense connections are able to support sustained reanalysis, and thus one would suspect the average level of morphological composition to be higher. Conversely the effect of hierarchical social structure, which emerges as human social networks grow, places a pressure for innovations that ease communication rather than being the product of repeated reanalysis. In these hierarchical structures, the previous reliance on morphological composition is replaced with more syntactic composition. This complex interaction between network structure and usage-based transmission provides the first mechanistic explanation consistent with both linguistic theory and observed history of natural language change.

8. Open data

All raw data and code to generate and analyze it are available at <https://gitlab.com/skagit/SNLLC> for public use.

Acknowledgments

We thank Shimon Edelman, Siew-Ann Cheong, and Randy LaPolla for commenting on a previous draft of this manuscript, and Stefan Frank and Matthew Spike for commenting on a partial presentation of this work at the 39th Annual Meeting of the Cognitive Science Society. This work was supported by Singapore Ministry of Education's Tier 1 grant #M4011320 to LO.

Note

1. Within the linguistics literature, extension is referred to as analogy, while reanalysis refers specifically to changes in grammatical structure, while an additional term, bleaching, is used to refer to changes in semantic structure. In this paper we highlight the more general means of novel usage with the term extension; much novelty that leads to language change has nothing to do with analogy as conventionally understood in the context of English grammar. Similarly, structural change happens at various levels of language use, and so we use the term reanalysis throughout to refer to structural change regardless of kind.

References

- Barabási, A.-L. (2014). *Network science*. Cambridge, UK: Cambridge University Press.
- Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286, 509–512.
- Batali, J. (1998) Computational simulations of the emergence of grammar. In J. R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approach to the evolution of language* (pp. 405–426). Cambridge, UK: Cambridge University Press.
- Bentz, C., & Winter, B. (2013). Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change*, 3, 1–27.
- Bromham, L., Hua, X., Fitzpatrick, T. G., & Greenhill, S. J. (2015). Rate of language evolution is affected by population size. *Proceedings of the National Academy of Sciences*, 112, 2097–2102.
- Bybee, J., Perkins, R., & Pagliuca, W. (1994). *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.
- Carneiro, R. L. (1967). On the relationship between size of population and complexity of social organization. *Southwestern Journal of Anthropology*, 23, 234–243.
- Chakrabarti, M., Heath, L., & Ramakrishnan, N. (2017). New methods to generate massive synthetic networks. *arXiv preprint arXiv:1705.08473*.
- Dale, R., & Lupyan, G. (2012). Understanding the origins of morphological diversity: The linguistic niche hypothesis. *Advances in Complex Systems*, 15, 1150017.
- Dryer, M. S., & Haspelmath, M. (Eds.) (2013). *WALS Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at <http://wals.info/>.
- Dunbar, R. (1998). The social brain hypothesis. *Brain*, 9, 178–190.
- Erdős, P., & Rényi, A. (1959). On randomgraphs, i. *Publicationes Mathematicae (Debrecen)*, 6, 290–297.
- Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, 32, 429–448.
- Fortescue, M. D. (2016). Polysynthesis: A diachronic and typological perspective. *Oxford Research Encyclopedia (linguistics)*.
- Gonçalves, B., Perra, N., & Vespignani, A. (2011). Modeling users' activity on twitter networks: Validation of dunbar's number. *PLoS ONE*, 6, e22656.
- Grünwald, P. D. (2007). *The minimum description length principle*. Cambridge, MA: MIT Press.
- Han, C.-h., Musolino, J., & Lidz, J. (2016). Endogenous sources of variation in language acquisition. *Proceedings of the National Academy of Sciences*, 113, 942–947.
- Heine, B. (2003). On degrammaticalization. *Amsterdam Studies in the Theory and History of Linguistic Science Series*, 4, 163–180.

- Herrera, C., & Zufria, P. J. (2011). Generating scale-free networks with adjustable clustering coefficient via random walks. In *Network Science Workshop (NSW), 2011 IEEE* (pp. 167–172). IEEE.
- Hickey, R. (2010). *The handbook of language contact*. Sussex, UK: John Wiley & Sons.
- Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization*. Cambridge, UK: Cambridge University Press.
- Jacques, G. (2012). From denominal derivation to incorporation. *Lingua*, *122*, 1207–1231.
- Jin, E. M., Girvan, M., & Newman, M. E. (2001). Structure of growing social networks. *Physical Review E*, *64*, 046132.
- Ke, J., Gong, T., & Wang, W. S. (2008). Language change and social networks. *Communications in Computational Physics*, *3*, 935–949.
- Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*, 10681–10686.
- Langley, P., & Stromsten, S. (2000). Learning context-free grammars with a simplicity bias. In *European Conference on machine learning* (pp. 220–228). Berlin: Springer.
- Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PloS ONE*, *5*, e8559.
- Lupyan, G., & Dale, R. (2016). Why are there different languages? The role of adaptation in linguistic diversity. *Trends in Cognitive Sciences*, *20*, 649–660.
- McCarty, C., Killworth, P. D., Bernard, H. R., Johnsen, E. C., & Shelley, G. A. (2001). Comparing two methods for estimating network size. *Human Organization*, *60*, 28–39.
- Narrog, H., & Heine, B. (2011). *The Oxford handbook of grammaticalization*. Oxford, UK: Oxford University Press.
- Nettle, D. (2012). Social scale and structural complexity in human languages. *Philosophical Transactions of the Royal Society B*, *367*, 1829–1836.
- Newman, M. E. (2003). The structure and function of complex networks. *SIAM Review*, *45*, 167–256.
- Newman, M. (2010). *Networks: An introduction*. Oxford, UK: Oxford University Press.
- Perfors, A., & Navarro, D. J. (2014). Language evolution can be shaped by the structure of the world. *Cognitive Science*, *38*, 775–793.
- Ravasz, E., & Barabási, A.-L. (2003). Hierarchical organization in complex networks. *Physical Review E*, *67*, 026112.
- Realí, F., Chater, N., & Christiansen, M. H. (2018). Simpler grammar, larger vocabulary: How population size affects language. *Proceedings of the Royal Society B*, *285*, 20172586.
- Ringe, D. (2008). *From Proto-Indo-European to Proto-Germanic*, vol. 1. Oxford, UK: Oxford University Press.
- Roberts, G., Lewandowski, J., & Galantucci, B. (2015). How communication changes when we cannot mime the world: Experimental evidence for the effect of iconicity on combinatoriality. *Cognition*, *141*, 52–66.
- Sebba, M. (1997). *Contact languages: Pidgins and creoles*. New York: Macmillan.
- Senghas, A., & Coppola, M. (2001). Children creating language: How nicaraguan sign language acquired a spatial grammar. *Psychological Science*, *12*, 323–328.
- Smith, K., Brighton, H., & Kirby, S. (2003). Complex systems in language evolution: The cultural emergence of compositional structure. *Advances in Complex Systems*, *6*, 537–558.
- Soffer, S. N., & Vazquez, A. (2005). Network clustering coefficient without degree-correlation biases. *Physical Review E*, *71*, 057101.
- Tria, F., Galantucci, B., & Loreto, V. (2012). Naming a structured world: A cultural route to duality of patterning. *PloS ONE*, *7*, e37744.
- Trudgill, P. (2011). *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford, UK: Oxford University Press.
- Varga, I. (2015). Scale-free network topologies with clustering similar to online social networks. In *Proceedings of the international conference on social modeling and simulation, plus econophysics colloquium 2014* (pp. 323–333). Cham, UK: Springer.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Fig. S1. Supplemental plots.