

Cornell University
College of Arts and Sciences

Modeling Syntactic Devices:
An Exploration of Language Evolution from Connectionist
and Memetic Perspectives.

Gary Lupyan

Spring, 2002

A thesis submitted in partial fulfillment of the
requirements for a baccalaureate degree in
Cognitive Science
with honors.

Abstract

This thesis introduces the subject of language evolution and provides insights into the evolution of several syntactic devices by modeling the acquisition of artificial languages by humans and neural networks. Previous work has shown that some linguistic universals can be learned by sequential-learning devices with no language-specific biases (e.g., Ellefson & Christiansen, 2000; Christiansen & Devlin, 1997; Van Everbroeck, 1999). Here, we conduct a series of connectionist simulations, as well as an artificial-grammar experiment to examine the ways in which case markings and word order may function as cues for the acquisition grammatical roles—“who did what to whom.” We hypothesize that languages that use rare devices to communicate “who did what to whom” are rare partially because they are more difficult to acquire. The results confirm our hypotheses and accommodate patterns of syntactic development across several different languages. Our results are consistent with the view of “language as an organism” (e.g., Christiansen, 1994). On this account, language universals may reflect non-linguistic, cognitive constraints on learning and processing of sequential structure, rather than constraints prescribed by an innate universal grammar. Languages that are easy to learn would proliferate, while those that are difficult to learn would die out, or never come into existence in a process of cultural evolution. I extend this evolutionary approach by framing language as a powerful replicator in its own right, and by investigating interactions between language and thought in a cognitive-linguistic process of self-stimulation.

Acknowledgments

A huge thank-you to Morten Christiansen, my thesis advisor, without whose time, advice, and support this work would not be possible. I would also like to thank Becky Pilon for putting up with me through the often grueling process, and for alpha-, and beta-testing the experimental stimuli. My gratitude to Shimon Edelman and Michael Owren who took their time to make invaluable suggestions for improving this manuscript. Thanks also to my mom for her support and presence at the 4th International Evolution of Language Conference. Finally, I am indebted to Sarah Burger and Margaret Collins for turning this thesis red with their superb editorial comments.

dedicated to Becky—22 going on 65

Part I: Background

Humans speak; other animals don't. This dichotomy sums up the fascination people have had with answering the questions: where did language come from, and why is it the way it is? The degree to which language makes us human cannot be overstated. It is a necessary pre-requisite for culture—which is impossible without a system that allows for the passing of knowledge from generation to generation. Without such a system each generation would essentially start off from scratch having to rely solely on its genetic endowments. Not only does language allow us to maintain and spread acquired knowledge, but language itself is a system shaped by culture and by the learning constraints of the human brain.

In recent years, evolution of language has become an increasingly popular subject of inquiry (e.g., Knight, Studdert-Kennedy, & Hurford, 2000; Lieberman, 2000; Calvin & Bickerton, 2000; Deacon, 1997; Pinker, 1994). Advances in computational modeling have often made inquiries into the evolution and acquisition of language quite fruitful.¹ In particular, computational neural network models have proven particularly successful in understanding the acquisition of grammatical relations (Morris et al., in press), the acquisition of past tense (Rumelhart & McClelland, 1986), reading and dyslexia (Plaut & Shallice, 1993), disambiguation of words and meanings (Cottrell, 1989), and linguistic change over time (Hare & Elman, 1995) among others.

This thesis will focus on one particular aspect of language evolution—the evolution of word order and case systems. I will first introduce some key issues in language evolution and provide some justification to approaching language as a cultural product. I will then describe three neural-network simulations we designed to test our main hypothesis: languages with rarer devices for signaling syntactic relationships (“who did what to whom”) are rare partially because they are more difficult to learn by general-purpose sequence learning devices. These simulations are supplemented by a human artificial-language experiment which provides preliminary support for the idea that rare syntactic devices are also more difficult to learn by human subjects.

* * *

In an introductory chapter to psycholinguistics, Ratner, Gleason, and Narasimhan (1998, p. 5) write without qualification that “no real reason exists why English should require the particular grammatical conventions it does.” More specifically, “English is considered to have a basic word order in which subjects precede verbs and objects follow verbs. [However] this tendency to put subjects before verbs in English is no more ‘logical’ than an insistence the verb should come first [as it does in Arabic].”

According to Ratner et al., grammatical rules are as arbitrary as the words we use to refer to entities (e.g., nothing about *apple* suggests a round, juicy fruit). This view is compatible with the notion pervading modern generative linguistics of the independence of syntax from the rest of cognition. This is the doctrine of syntax autonomy (Chomsky, 1957).² In this view, syntax is processed separately from contextual and semantic information. Because syntax is presumed to be acquired through an innate, domain-specific device operating within the constraints of the so-called universal grammar (UG)—a system of rules and parameters that allow a child to acquire any human language (e.g., Chomsky 1988). Working in this framework, many have found it mystifying that the various rules and conventions occurring in languages around the world are by no means equally represented. If every rule is equally learnable through UG, why should some be more common than others?

In a widely-cited volume on cultural evolution, Cavalli-Sforza and Feldman (1981, p. 357) ask why do some languages use “elaborate declensions of nouns and conjugations of verbs” while others use “syntactic devices including affixes or prepositions.” They additionally ask “whether the sentence structure subject-verb-object (as in English) is preferable to that used, for instance, in Japanese: subject-object-verb?” Their conclusion: “the adaptive value of each system is unknown.”

¹ At least more fruitful than past research. In 1866, the Linguistic Society of Paris issued a ban on papers dealing with the origins of language due to the abundance of wildly speculative, and whimsical theories (see Aitchison, 1996).

² Chomsky has been more recently quoted saying, “It’s a logical impossibility for [John] Searle, or anyone, to differ with my thesis of the ‘autonomy of syntax,’ because I’ve never held any such thesis. There is a very large ‘debate’ about it, with many people attacking the thesis (but without telling us what it is) and no one defending it, surely not

The assumption that grammatical rules are arbitrary is clearly a non-answer in light of the strong constraints that operate on languages. We present data that suggest that the asymmetries seen in word-order and case patterns around the world arise, in part, from the learnability constraints of different syntactic devices. Specifically, languages that are more consistent in their use of syntactic cues are more learnable by sequence-learning devices (artificial neural networks) as well as by the human brain.

I will conclude by providing possible answers to two fundamental questions. The first is what drives linguistic change. The second concerns the seeming chicken-and-egg problem of language and cognition. We only need complex language if we have complex thoughts (Premack, 1986: “if we gave syntax to chickens, they’d have nothing to say!”). At the same time, complex language facilitates and is perhaps a pre-requisite to complex thought (Bickerton, 2002; see Dennett, 1991 for extensive discussion). I will approach this question by expanding on Daniel Dennett’s (1991) notion of language as a “cognitive autostimulator.”

A Brief Introduction to Evolution of Language

In 1986, David Premack penned the oft-quoted sentence: “language is an embarrassment for evolutionary theory” (p. 132). This statement was inspired by two observations: (1) how does a communication system that is qualitatively different from all others (Deacon, 1997) evolve through necessarily gradual natural selection? (2) How do we account for a human system of linguistic communication that appears to be vastly more powerful than one can account for in terms of selective fitness? I will attempt to answer the second question towards the end of this thesis by invoking ideas from cultural evolution (Dawkins, 1989; Cavalli-Sforza & Feldman, 1981; Blackmore, 1997).

In response to the first question, proponents of the view of language as a biological adaptation offer the less-than satisfactory explanation that the missing links

me, because I have no idea what it is (31 Mar. 1995)” (see Plato’s Problem, Orwell’s Problem, and Life in the

between extant apes which have so much trouble learning even simple languages (Deacon, 1997; Pinker, 1994), and humans who are so adept in acquiring languages—these ancestors bearing “half a language”³— are simply extinct. Just because chimpanzees are our closest extant relatives does not mean we should expect them to bear traces of linguistic homologues (Pinker, 1994). Viewing language as a culturally evolved system based on non-linguistic mechanisms potentially solves this problem. While there are no clear biological homologues to a “language module,”⁴ there are obvious homologues to sequence-learning devices in non-human animals. As will become apparent shortly, the learning and use of language is closely linked to learning and executing sequences (e.g., Lieberman, 2000). This should come as no surprise; after all, language is a strongly sequenced system in which order matters. In addition, higher-order relations and hierarchies are essential to language. While many non-human animals are capable of learning simple sequences, and non-human primates can learn fairly complex sequences, only humans seem capable of the type of hierarchical representation necessary for language (see Conway & Christiansen, 2001 for review of sequential learning in primates). Sequence-learning is thus an excellent candidate for what makes humans the only animals to be capable of learning language.

In this view, language emerged from mechanisms evolved for non-linguistic, sequential-learning tasks. The impetus for an ever more complicated ability to process and conceptualize patterns may have come from tool-making (e.g., Christiansen & Devlin, 1997). Some have suggested that the freeing-up of arms by bipedalism produced strong evolutionary pressures favoring ever more complex sequencing of motor actions (e.g., Lieberman, 2000). There are many interesting possibilities, but they are not the focus of this thesis. We will assume that during the formation of early languages, humans already had most of the “neural hardware” that made symbolic

Spotlight Located At: <http://mitpress2.mit.edu/e-books/chomsky/4/15.html>).

³ Or half a Universal Grammar (Pinker & Bloom, 1990).

⁴ Work on mirror neurons (e.g., Arbib, 2002), that is, neurons that fire when both performing an action and observing the action being performed, promises to be the missing link between non-human and human primates in the evolution of language. Mirror neurons that fire both when performing and observing grasping actions have been found in monkeys in the premotor area F5, a homologue to Broca’s area in humans. However, Arbib himself has admitted that mirror neurons are likely to be very common among mammals and are probably widely distributed throughout the brain. For instance, it is likely that schools of fish use some type of mirror-neuron system for synchronized swimming—if you see a fish in front of you turn left, turn left yourself. Mirror neurons are likely to be very basic neurological structures and therefore provide poor candidates for direct language pre-cursors.

thought, compositionality and formation of symbolic hierarchies possible. If such an infant could be adopted into a modern family, she could learn any modern language.

* * *

An issue that warrants discussion is the evolution of speech. The human vocal tract is unique among primates in that food and water have to pass over the tracheal opening on the way to the esophagus, significantly magnifying risks of choking. This odd anatomy greatly facilitates speech by expanding the vowel space through increased mobility of the larynx. But how could a lowered larynx be selected for, given the maladaptive side-effects of such a structure, if speech weren't already trait linked to reproductive success? Paradoxically, without a lowered larynx, speech would be muddled and labored. Lieberman has argued (2000, 2002) that the only redeeming quality of the lowered larynx is the ability to pronounce the vowel [i] (as in tree)—an adaptation seemingly restricted to language. This seems difficult to dispute until we consider the possibility opened by Owren and Rendall (2001). Many animals produce two basic types of signals—screams and shrieks, as well as formant-based (i.e., vowel-like) vocalizations such as gruffs and sonants. While the broadband nature of the former masks any individual differences, formant-based vocalizations have acoustics strongly shaped by the caller's vocal tract and can be effectively used for individual identification. Primates are able to discriminate formants on a level comparable to humans, and their role in identification appears to be consistent with primate vocal behavior (Owren & Rendall, 2001). When combined with the social group theories expressed by Dunbar (2002), perhaps we can posit a nonlinguistic reason for selecting for better production and perception of formants—namely, for keeping track of cheaters and reciprocators. This suggests that in addition to a non-linguistic physical evolution of a sequential-learning device that has come to underlie language, there may have been non-linguistic pressures to evolve greater control of the larynx.

Another line of evidence suggests that language was originally based on gestures (e.g., Goldin-Meadow, 1999; Corballis, 2002a). In time, with the improvement of the vocal apparatus and voluntary motor control of the larynx and tongue, gestures

became decoupled from vocal signals—the latter became sufficiently rich to be used alone (Corballis, 2002b). Humans still use gestures to supplement speech, and people can readily learn sign languages, now known to be as rich and full as their vocal analogs. Additionally, congenitally blind children spontaneously gesture to others, and aphasias impair both signed and vocal languages in very similar ways (Bellugi, 1983). Regardless of whether a gestural or vocal system came first, converging evidence suggests that both systems are mediated by common neural structures.

Furthermore, language seems to be integrally connected to the rest of cognition—in particular, to the learning of sequences. For instance, speech deficits characteristic to the breakdown in motor control and sequencing experienced by Parkinson’s patients are not limited to speech, but pervade the ordering of thought itself (Lieberman, 2000). Similarly, aphasics do not just have problems verbalizing ideas, but exhibit “loss of the ‘abstract’ attitude” as well as deficits usually associated with frontal lobe damage: “[difficulties] planning activities, shifting strategies, formulating abstract categories, and thinking symbolically” (Lieberman, 2000 p. 99).

The few documented cases of alleged dissociations between language and the rest of cognition include family KE—a British family containing a large number of individuals with a recently identified genetic disorder (Lai et al., 2001) rendering them virtually incomprehensible while seemingly sparing nonverbal intelligence (Gopnik & Crago, 1991). More exhaustive testing by Vargha Khadem et al. (1995) revealed that in addition to speech, these individuals had severe problems ordering non-linguistic motor movements.

While these failures to find linguistic dissociations may be surprising to linguists and others who view language as a system mysteriously separate from the rest of cognition, this is not at all surprising if we consider that language production systems—whether vocal or gestural, as well as language perception—vocal, gestural, visual (reading), tactile (reading Braille), or gesturo-tactile (as in the case of Helen Keller) all involve sequential and hierarchical processing of information. It often escapes notice that the two dominant language modalities: vocal and gestural, are actually one and the same—both involve fine motor movements—the former, of the tongue and larynx; the latter of hands, arms, and face.

Part II: Experiments

Introduction

One of the most difficult tasks children are faced with when acquiring language is mapping a sequence of words onto an interpretation of what that sequence is supposed to mean. In order for the child to understand a sentence, she needs to determine the grammatical roles of the individual words so that she can work out who did what to whom. Although children appear to bring powerful statistical learning mechanisms to bear on the acquisition tasks (e.g., Saffran, Aslin & Newport, 1996), the existence of linguistic universals common across radically different languages (Greenberg, 1963) points to the presence of innate constraints on such learning. Without such constraints, it becomes difficult to explain why there are few, if any, Object-Subject-Verb (OSV) languages (van Everbroeck, 1999) even though in principle such a language appears to be as good as any other. Here, we propose that these constraints may arise from non-linguistic limitations on the sequential learning of statistical structure, and examine how this perspective may shed light on how children learn to map the words in sentences onto their appropriate grammatical roles. There are two major ways in which languages signal syntactic relationships and grammatical roles—word order (WO), and case markings. In a strict WO language, declarative sentences follow a predetermined order—Subject-Verb-Object (SVO) in English. It is the occurrence of the subject in the first position, and the object in the second, that allows the hearer to comprehend who did what to whom. In contrast, languages such as Russian or Japanese allow multiple word orders and rely on case markings to disambiguate subjects from objects. For instance, *Masha lubit Petyoo* (SVO), *Petyoo lubit Masha* (OVS), and *Lubit Petyoo Masha* (VOS) are all grammatical in Russian and all mean “*Mary loves Peter*” (albeit with different emphases on the constituents). This semantic invariance is possible due to the nominative *-a* and accusative *-oo* case markers.

While long-standing theories describe acquisition of language through an innate language acquisition device (e.g., Pinker, 1995), an alternative approach that is gaining

ground is the adaptation of linguistic structures to the human brain rather than vice versa (e.g., Christiansen, 1994; Kirby, 1998). On this account, language universals may reflect non-linguistic cognitive constraints on learning and processing of sequential structure rather than constraints prescribed by an innate universal grammar. Previous work has shown that sequential-learning devices with no language-specific biases are better able to learn more universal aspects of language as compared to aspects encountered in rare languages (e.g., Christiansen & Devlin, 1997; Ellefson & Christiansen, 2000; Van Everbroeck, 1999, 2001).

Here, we examine the ways in which case markings and word order may function as cues for a sequential learning device acquiring syntactic structure. In simulation 1, we model different word orders and hypothesize that a sequential learning device should have an easier time learning typologically common languages than the more rare ones. We expand on this idea in simulation 2 by studying the performance of networks trained on languages of varying degrees of case markings and flexibility. Finally, in simulation 3, we establish that our trained networks are able to mimic syntactic performance of children speaking English, Italian, Turkish, and Serbo-Croatian (Slobin & Bever, 1982).

Acquisition of Word Order

Generative linguists have long relied on parameter setting to explain how children acquire the distinct patterns of their native language. For instance, it has been assumed that the way a child knows to generate SVO and not SOV English sentences is through the setting of a VO/OV parameter (Neeleman, 1994). This is unsatisfactory because this account does not account for many observed correlations. For instance, OV languages typically have flexible word orders (Koster, 1999), while most VO languages have strict word order. More generally, parameter theory has been largely unable to account for the asymmetries and patterns in the distribution of world languages. Why, for instance, are the most common word orders overwhelmingly SOV, SVO, and VSO (Greenberg 1963: Universal 1)? Why do verb-final languages almost always have a case system (Greenberg 1963: Universal 41)? And even more

fundamentally, why do case languages have flexible word orders to begin with? It is our position that these observations can be at least partially accounted for by examining the learnability of languages from the viewpoint of sequential learning.

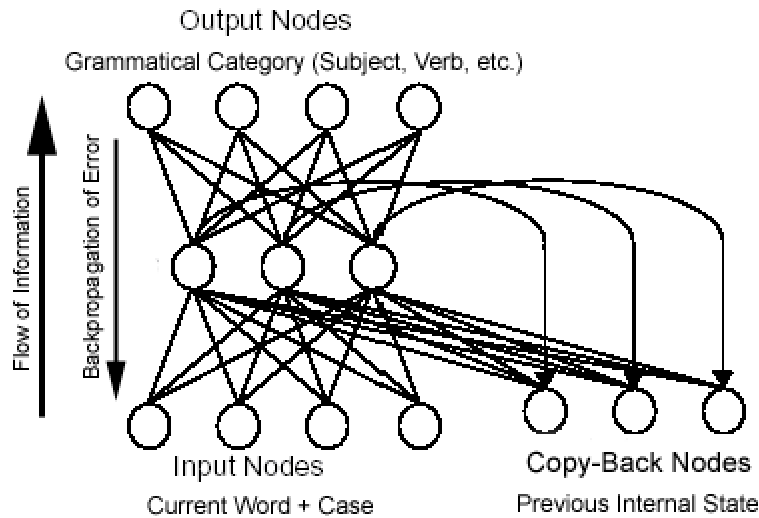
Generative linguistics also leaves largely unexplained the process children use to actually set the parameters. With regard to word orders, an explanation espoused by Pinker (e.g., 1995) involves the so-called Subset Principle. According to the Subset Principle, children take the most conservative strategy and so, by default, assume a fixed order. Alternative word orders are only accepted if a child is exposed to these orders, at which time a free word order parameter gets switched on. Under this assumption, FWO languages are predicted to be more difficult to learn. Although the idea that all languages are initially approached as having strict word-order (SWO) was popular in the sixties and seventies (Slobin, 1966), Slobin and Bever (1982) conclude that the primacy assigned to word order was unduly influenced by languages such as English.

There is ample evidence that children learning a strict word-order language such as English never leap to the conclusion that it is a free-word order language (Pinker, 1995). While Pinker has used this evidence to reinforce parameter-setting, stating that the reason children never leap to such conclusions is because a word-order parameter has been set, we suggest an alternative explanation. Simply stated, children learning English generally do not produce non-SVO sentences because non-SVO sentences are incomprehensible in English. In the absence of case markings, *Kicked John Bill* is ambiguous as to who did the kicking. Children learning English use the statistical properties of the language to learn that word order is a reliable cue to syntactic relationships. Children learning a case-based language, such as Russian, make a similar observation about case markings. This view obviates the need for a default strategy. What is important is that there exist some set of cues to indicate syntactic relationships—there is nothing inherently special about word order or case markings. In short, we posit that a major reason for the observable asymmetries among the world's languages is that certain patterns make a language more easily learnable by a general sequential-learning device, ensuring the proliferation of such a language in the human population.

Simulation 1: Learnability of Case and Word Order

If frequency of certain WOs is correlated with the learnability of languages that employ them, typologically rare languages should be more difficult to learn by a sequence-learning device than the more common languages. To test this prediction, we trained simple recurrent networks on a total of 14 artificial grammars reflecting the 6 possible strict word orders (SWO) and a flexible word order (FWO), with or without the presence of case markings.

Simple recurrent networks (SRN: Elman, 1990) are based on simpler back-propagation networks. Both belong to a class of error-driven networks, meaning the produced output is compared with the expected output. The computed difference between the two outputs is the so-called error, and is then used to adjust the weights between the output, hidden, and input layers of the network. Because the changes during each presentation are very small, such networks require a large number of trials to converge. Simple recurrent networks expand this concept by adding a context layer (also called copy-back) with connections back into the hidden layer (see Fig. 1). These recurrent connections provide the network with a simple form of short-term memory and make it particularly well-suited for learning sequences. After the presentation of each stimulus (here, word), the current hidden layer is copied into the context layer. During the presentation of the next word, the context layer is copied *back* into the hidden layer along with the current input. It is this copying that provides the network with the context in which to process the current stimulus.



Adapted from Plunkett & Elman (1997).

Figure 1: A schematic diagram of a simple recurrent network (SRN).

Method

Networks

Ten SRNs were used in each condition. The networks were initialized with random weights in the interval $[-0.1, 0]$. It was found that the slightly inhibitory starting weights provided better performance across the board. A similar conclusion was reached by van Everbroeck (2001). Each input to the networks consisted of a distributed representation of a word, spliced with a case marker. Words were represented by 20-unit randomly generated bit-vectors. Although some vectors were bound to be close in the representation space, random assignment to words assured that any such interaction would not bias the results. Having words represented by random vectors may seem odd considering the complex phonology that underlies human languages, but for present purposes such a representation seemed to work just as well as phonological (e.g., van Everbroeck, 2001), while dramatically decreasing

training time. Case markings (nominative, accusative, dative, and genitive) were represented by a 4-bit vector appended to the word vector. This made for a total of 24 input units. There were 7 output units, corresponding to grammatical roles the network was supposed to predict: subject, direct object, indirect object, genitive noun, verb, or end-of-sentence. In all simulations, the learning rate was set to 0.1, and momentum to 0.01. Each SRN had 30 hidden units and 30 context units.

Materials

The lexicon contained 300 nouns and 100 verbs. This noun-to-verb ratio is generally consistent with human languages (e.g., British National Corpus). The verbs were evenly divided into intransitive, transitive, and ditransitive categories. As illustrated in Table 1, each grammar included three types of sentences: intransitive, transitive, and ditransitive. A sentence consisted of noun phrases (NP) and one of three verb classes. Twenty-five percent of noun phrases contained a noun in the genitive form (e.g., *John's brother*). The simplest sentence generated by such a grammar was a simple intransitive: e.g., *John walks*. The most complex sentence contained 7 words: *Mary's friend gave John's brother Peter's key*. A fully flexible grammar was identical to the strict WOs except the order within each element was randomly varied from sentence to sentence. In an effort to model the languages as naturalistically as possible, we modeled genitives based on Greenberg's (1963) universal 2: in typically prepositional languages (SVO and VSO) genitives generally follow the governing noun, while the reverse is true in postpositional languages (SOV). We modeled the remaining three word orders with genitives following the noun. We also added a genitive case-marking to SWO-no case languages. Without this, it was impossible for the networks to discern governing nouns from genitives. This addition is motivated by the observation that even normally case-less languages have some form of genitive case markings (e.g., in English: *Mary's house*) (van Everbroeck, 2001).

Table 1: A sample SOV Grammar Used to Generate the Corpora

S → Intransitive [0.35] Transitive [0.35] Ditransitive [0.3]
NP → Noun, Noun-Gen [0.25]
Intransitive → NP-Nom, Intransitive-Verb
Transitive → NP-Nom, NP-Acc, Transitive-Verb
Ditransitive → NP-Nom, NP-Acc, NP-Dat, Ditransitive-Verb

Procedure

We used a crossover design of 7 word orders (6 strict and 1 flexible), by two case conditions (with or without case) resulting in 14 training corpora. For each condition, we generated 3,000 random sentences of the appropriate order. Such a corpus occupies a very small part of the possible sentences that can be generated by the corpus. For instance, 9 million different sentences are possible for a transitive SOV configuration (300 x 300 x 100).

The networks were trained for 100,000 sweeps (input/output pairs), corresponding to about 7 passes through the corpus. During each training sweep, the network was presented with a word, and depending on the condition, a case marking. A corpus of 200 novel sentences was created for testing. In the testing corpus, 50% of words were completely new—ones to which the network has never been exposed. Performance was measured by assessing the network's ability to map a given word to its correct grammatical role. During testing, the network's highest-activated output unit was compared to the expected output. If the units matched, the word was marked as being correctly mapped.

It may seem that providing the networks with direct mapping from word to grammatical category is unrealistic and not ecologically valid. After all, it has long been recognized that children are not given sufficient ostensive cues to syntactic relationships and word meanings—this is the so-called “gavagai problem” (Quine, 1960).⁵ No one explains to the child after each encountered sentence that word *A* was the “do-er” and word *B* the “do-ee”. However, Tomasello and colleagues have shown that children are

⁵ Suppose you came across a tribe who used a word *gavagai* whenever they saw a rabbit running across a field. From just these encounters it would be impossible to determine if *gavagai* referred to the whole rabbit, just its ears, or even the act of running.

able to use pragmatic cues such as eye gaze to help figure out which object is being referred to (Tomasello & Akhtar, 1995). Twenty-four month olds show understanding of adult intentions in inferring meanings of novel verbs (Tomasello & Barton, 1994), and 18-month old children are able to learn new words in non-ostensive contexts (Tomasello, Strosberg & Akhtar, 1996). Such use of pragmatic cues enables children to map words onto meanings and correctly infer who did what to whom. Considering that our networks live in a purely linguistic world, without access to the types of pragmatic cues used by children, direct mapping of words onto syntactic roles seems like a reasonable learning paradigm.

Results

All networks trained in the case condition were able to map 100% of the words to the correct categories. When case was not available, the network performance roughly corresponded to attested language frequencies (Table 2). Two WOs obtained nearly perfect performance: SVO and VSO (99%). The results for the other word orders appear in Table 2. The likely reason SOV-no case grammars did not achieve perfect accuracy is because they contained two unmarked nouns prior to the verb. Since the networks learn to map different types of verbs to different argument constructions, verb-final grammars are at a disadvantage—in these grammars the grammatical role that provides the most information about what is to come is received last (van Everbroeck, 2001). This finding corresponds to Greenberg's (1963) observation that verb-final languages almost always have a case system, presumably to counteract this statistical disadvantage. In contrast, many SVO and VSO do without case markings (e.g., English, Welsh) (Newmeyer, 2000). We should also note that even though FWO-no case languages perform poorly, their performance is consistently above chance. This can be explained by the networks' learning to map familiar verbs to intransitive, transitive or ditransitive word schemas. When word familiarity was eliminated (only novel words were used), performance dropped to chance.

From the network's perspective, the reason for the poor performance of some languages can be explained by inconsistencies in constructions across sentence types. For instance, when the network is presented with the first word of an OSV sentence, it is

not clear whether that first word is the subject of the intransitive (SV) or the object of a transitive sentence (OVS). The poor performance of VOS is similarly due to the intervening indirect subject in ditransitive sentences creating inconsistencies between the order encountered in transitives and ditransitives.

Table 2: Network performance and Language Distributions		
Word Order	Words Correct – No Case Condition (%)	Attested frequency in world languages (%) ⁶
SOV	90	51 (most employ cases)
VOS	85	8
OVS	80	0.75
OSV	74	0.25
Flex	65	0 (all employ cases)

All performance differences are highly significant at $p < 0.0001$

These simulation results confirm the idea that FWO-case languages are no more difficult to learn than common SWO-no case languages such as SVO and VSO. The difficulty associated with learning a FWO or a rare SWO language without case markings is underscored by typological evidence, suggesting that such languages use case markings to signal grammatical relationships (Payne, 1992).

Simulation 2: The Impact of Case on Word Order Flexibility

In natural languages, case markings are not wholly deterministic. For instance, Slavic languages such as Russian and Serbo-Croatian contain a number of nouns which, perhaps for historical reasons, do not take case markings. Additionally, because these markings often take the form of suffixes, they change the phonology of words. This results in potential phonological ambiguity. For instance, in Russian *stali* is either the genitive form of *steel* or a conjugated verb meaning *we stopped*. By examining the effects of varying cases on different word orders, we hoped to show that (1) even probabilistic case markings improve performance for FWO languages, and (2) case markings do not improve performance for languages that already rely on WO.

⁶ Adapted from van Everbroeck, 1999

Method

Networks

Ten SRNs were used for each condition. The initial conditions and training details were the same as in simulation 1.

Materials

We generated five artificial grammars varying in the salience of case markings, from only genitive markings to full case markings. A grammar with case marked 50% of the time corresponded to a language in which 50% of case markings were possibly phonologically ambiguous, or a language in which certain nouns did not take on case markings. Five more grammars varied on strictness of word order, from a completely flexible order to a completely strict one (SVO). The word orders approximated distributions found in natural languages (Italian and Turkish: Slobin & Bever 1982; Polish: Jacennik & Dryer 1992). The two conditions were crossed, for a 5x5 matrix. As in simulation 1, 3000 sentences were generated for each condition.

Procedure

Each group of networks had case cues added to the sentences based on case condition. The testing proceeded as in simulation 1.

Results

As expected, SWO languages inspired by English and Italian were little benefited by case (Fig 2). In contrast, the probabilistic addition of case markings to FWO languages consistently improved performance. The slightly lower performance of Italian was due to it having a more flexible word order than English (see Table 3). To compensate for possible ambiguities, Italian relies heavily on prosodic and contextual information (Slobin & Bever, 1982), which was not available to our networks.

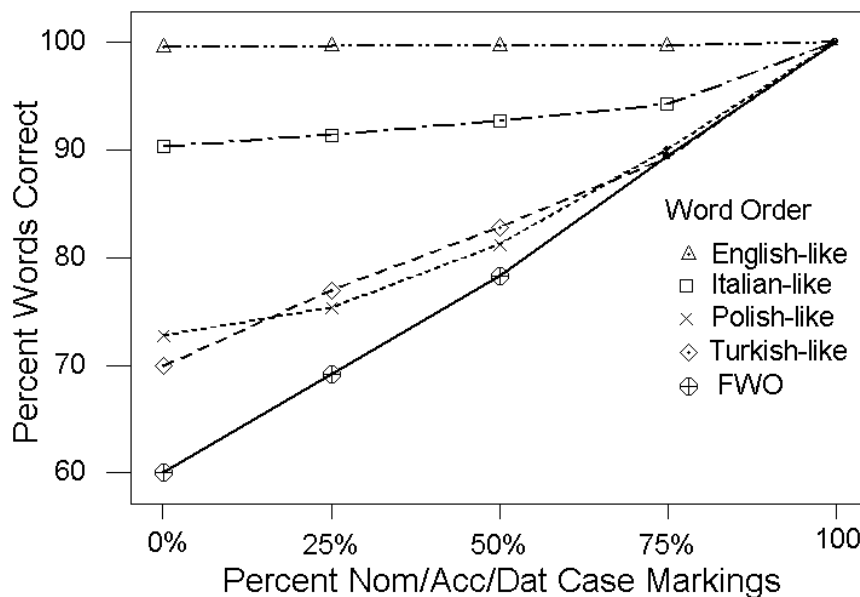


Figure 2: Network performance in simulation 2 for increasing degrees of case markings as a function of word order.

Simulation 3: Interactions between Case and Word Order Flexibility in Development

In this simulation, we demonstrate that networks trained on corpora similar to those used in simulation 2 are able to mimic syntactic performance of children learning English, Italian, Turkish, and Serbo-Croatian. Slobin and Bever (1982) tested 48 children in 8 age groups (24-52 months). Each child was tested on their ability to demonstrate familiar actions (e.g., *scratch*, *bump*, *pick up*) using familiar toy animals after hearing a transitive sentence in their native language. The authors hypothesized that Turkish, English, and Italian-speaking children would have the easiest time due to the consistent, unambiguous case markings available in the case of Turkish, and the consistent word-order information available in English and Italian. Children acquiring Serbo-Croatian would have a more difficult time because of its more ambiguous case markings, requiring them to pay attention to word-order as well as cases.

Method

Networks

The networks and training details were identical to those used in simulation 2. We used 12 SRNs in each condition, mirroring the number of subjects used by Slobin and Bever (1982).

Materials

We created types of grammars motivated by the four languages used in the study. English was modeled as being 100% SVO, and having only genitive case markings. The word orders for the remaining languages were modeled based on the data provided by Slobin and Bever's (1982) corpus of adult speech, reflecting the linguistic input available the children.

Table 3: Word Order Distributions for Simulation 3		
Language	Admissible Word Orders	Case Structure
English	100% SVO	Genitive only
Italian	82% SVO, 2% SOV, 11% VSO, 5% OVS	Genitive only
Serbo-Croatian	55% SVO, 16% SOV, 16% VSO, 3% VOS, 2% OVS, 8% OSV	Full for non SVO. For SVO: 55% nom, 55% acc, 100% dat, 70% gen
Turkish	48% SOV, 25% SVO, 13% OVS, 8% OSV, 6%VSO	Full

Although Turkish does not have an explicit nominative case, it was found that such a marker was necessary in this simulation. In the absence of semantic information and case markings, the networks must rely on the syntactic position of a word to correctly identify its category. However, in a relatively FWO language such as Turkish, this

information is ambiguous. Without a nominative case, both verbs and subjects are unmarked, and the network naturally has trouble telling them apart. In contrast to these networks, children rely on semantic information, in addition to syntax, to tell apart verbs and nouns. In other words, a Turkish child knowing the meanings of “dog” and “sniff”, will not confuse the two even when “dog” is an unmarked agent in the sentence.

Table 4: Percentage correct performance for grammatical sentences in a given language in the Slobin and Bever (1982) study.				
Language				
Age	English	Italian	Serbo-Croatian	Turkish
24-28	58	66	61	79
32-36	75	78	58	80
40-44	88	85	69	82
48-52	92	90	79	87

Serbo-Croatian has all four of the cases we were modeling, but only masculine and feminine nouns take on accusative and nominative markings. Sentences containing one or more neuter nouns are typically ordered as SVO. We did not have data on the proportion of neuter nouns in Serbo-Croatian or the percentage of SVO sentences containing such nouns. It was estimated that about 55% of SVO sentences would contain one such noun, therefore case markings were deleted from 55% of nouns in SVO sentences. Serbo-Croatian neuter nouns do have dative case-markings—hence, the datives are marked 100% of the time. However, plural neuter nouns are not declined in genitive constructions. If plural genitive nouns are used an estimated 30% of the time, then 70% of SVO sentences will have genitive case markers.

Procedure

Training proceeded as in simulation 1. The extent of training was varied for networks corresponding to different age groups. Testing was done following the procedure employed by Slobin and Bever (1982). We used transitive sentences using only words which the networks have seen during training. Performance was quantified by measuring the percentage of subjects and objects the network identified correctly, and averaging the data with the overall percentage of words correctly identified.

Results

The networks' performance (Fig. 3) closely matched Slobin and Bever's (1982) data (Table 4). As predicted, networks trained and tested on Turkish had the easiest time predicting the subjects and objects. Networks trained on Serbo-Croatian had the most difficult time, highlighting the higher processing-cost associated with having to pay attention to WO and case markings. Performance on Italian was slightly worse than on English, reflecting the more flexible WO of Italian. It is predicted that with the addition of prosodic and semantic cues, the performance of Italian would more closely parallel that of a fully SWO language such as English.

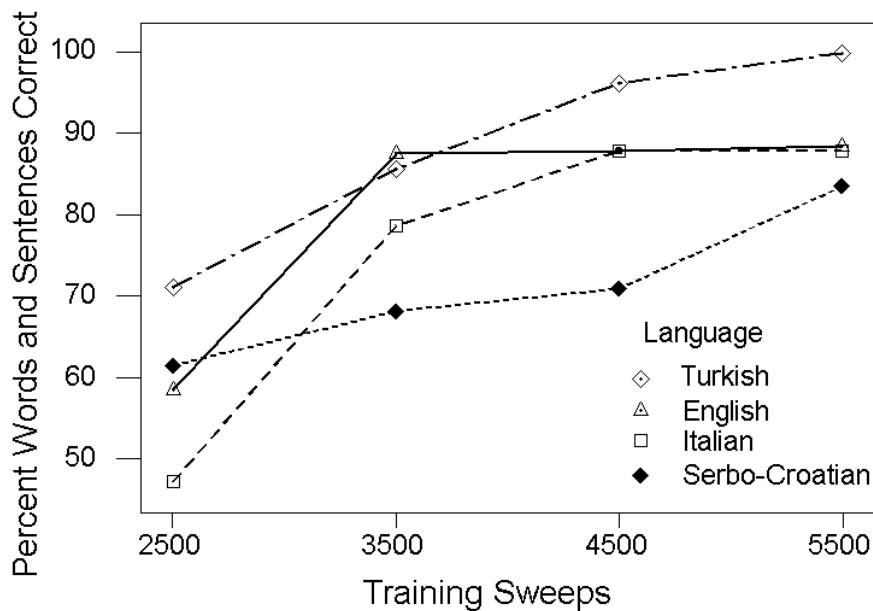


Figure 3: The pattern of performance across training for Turkish, English, Italian, and Serbo-Croatian in simulation 3

Conclusion

Our findings confirm that learnability of languages may be a major factor in the frequency of certain language types. In the view of language as an organism, languages

that are easily learnable by the human sequential-learning device proliferate, while languages not easily learnable die out or never come into existence. Our simulations suggest that all that is needed to learn syntactic relations is a reliable cue: case or word order; neither needs to be primary. No parameter-setting or subset principle is needed to account for the data. These results also provide added support for a connectionist approach to studying acquisition and evolution of language (Christiansen & Dale, 2000).

One question that remains is why there do not appear to be languages that actively use both cases and word order for syntactic disambiguation. We ran a simulation in which networks were trained on an SVO-Case language and then tested on alternative SOV and VSO grammars. Interestingly, even though the networks had never seen these WOs, they were able to identify the syntactic relations in the sentences with higher than chance frequency. This was not true for networks trained on SVO without case. Since case allows, perhaps even encourages, alternative word orders (e.g. for signaling emphasis), a hypothetical case-based SWO language will likely drift to a more FWO language with time.

The simulations described here have several notable limitations. The sentences used for training were admittedly simple. Although simple intransitive, transitive, and ditransitive sentences are very frequent in speech, natural languages are rife with more complex structures such as relative clauses and embedding. Offsetting the simple grammars were the limited cues available to the networks, which relied solely on distributional information of grammatical categories. In contrast, children routinely use semantics and prosodic cues, and even more subtle cues such as differential word length of nouns and verbs (Cassidy & Kelly, 1991). It is therefore quite remarkable that relying only on word order or case, the performance of the networks was near-perfect for common language types.

While simulation 3 shows that the networks used here are able to mimic human data, we went on to conduct an artificial grammar experiment on human subjects to corroborate the simulation results.

Artificial Grammar Learning Experiment

Introduction

After obtaining positive results from the simulations, we set out to test our word-order learnability hypothesis by exposing college students to several artificial grammars. The hope was that human data would confirm the results we obtained in the connectionist models, specifically, that FWO caseless languages would be the hardest to master due to the absence of reliable cues to syntactic relationships. We further hypothesized that FWO-Case and SWO-No case languages should be about equally easy to learn. SWO-Case languages may be easy to learn (as was the case in the simulations), or they may be more difficult than SWO-Case languages reflecting the cost of processing redundant information (Kirby, 1997). The artificial-grammar learning procedure we used is a derivative of the one used by McDonald and Plauché (1997), Ellefson and Christiansen (2000), and Conway and Christiansen (2002). These studies (esp. McDonald and Plauché) suggest that human subjects can use order of elements and morphological cues to deduce simple rules.

Subjects

Twenty-eight Cornell undergraduates participated in this experiment. For their participation they received psychology course credit or \$5.

Procedure

Subjects were randomly assigned to one of four language conditions: SWO with case (*SWO-C*), SWO without case (*SWO-NC*), FWO with case (*FWO-C*) and FWO without case (*FWO-NC*). The experiment consisted of two parts: *training* and *testing*. During the former, subjects heard 4 blocks of 2-4word artificial sentences (116 total). The order of the sentences in each block was randomized. After each sentence, the subject was asked to repeat it. The experimenter then marked the sentence as correct

or incorrect. The criteria for incorrect markings included omitted words, incorrect ordering of words, or cases in which the suffix of the word was clearly incorrect. Pilot data showed that repetition of these sentences was considerably more difficult than we anticipated. The instructions for the first part therefore included some words of encouragement: “We realize that the first part may be quite difficult. Don’t worry if you have trouble pronouncing some of the words. Just do the best you can.”

The second part, *testing*, consisted of 40 novel sentences. Twenty of these were grammatical and 20 ungrammatical. The subjects’ task was to determine which ones were which by pressing *Yes* for grammatical and *No* for ungrammatical on a button box. The performance measure we used was percent correct out of 40.

Materials

Four artificial languages were constructed, corresponding to each condition. Each artificial language consisted of three verbs (1 intransitive—*paazi*, 1 transitive—*tila*, 1 ditransitive—*lami*) and four nouns (*voopoo*, *zekkaa*, *tuzee*, *soogee*). The bisyllabic CVCV pattern was chosen for the ease of adding case markings by replacing the final vowel with a vowel case-marking in a deterministic fashion. This allowed us to maintain the length of words constant across the case/no case conditions.

The words were synthesized using ByteCool[®] TextSound’s⁷ “Mary” voice set to 145 words per minute at 169 Hz. The sentences were presented with brief pauses between words to prevent co-articulation effects. These speech parameters were set by trial and error to maximize intelligibility of the stimuli. The words themselves (e.g., *lami*, *soogee*, *zekka*) were chosen based on two criteria. Each word was selected to be unfamiliar to English speakers. If some words were more familiar than others, they may have been easier to remember and could have skewed the results. Several subjects thought *paazi* sounded a bit like *pause-y* but overall responses on a post-study questionnaire assured us that none of the words were familiar to subjects. When asked to write down the words they remembered after the experiment, none were more likely to be remembered than others. The second criterion was intelligibility: the words were

designed to be easily discriminable from one another as pronounced by Mary, the synthesized voice. Case markings were represented by vowels appended to the CVC root of the noun (nominative: -aa, accusative: -oo, and dative: -ee) e.g., tuzaa (nom.), tuzoo (acc.), tuzee (dat.). These suffixes were co-articulated with the preceding syllable.

Sentence Structure

As in the simulations, the sentence corpus consisted of three types of sentences: intransitive, transitive, and ditransitive. To limit the sentence length, we decided to not include genitives in this experiment, and so the sentences ranged from 2 to 4 words. Strict word order sentences were ordered as VSO reflecting a word order that frequently occurs without case markings, and one that is generally unfamiliar to English speakers. Free word order sentences were evenly divided between three word orders that are most typologically prevalent: VSO, SVO, and SOV.

Twenty-nine sentences were generated for each language. This number reflects the limitation in combinatorics of using 3 verbs and 4 nouns with which only 64 ditransitives ($1 \times 4 \times 4 \times 4$), 16 transitive ($1 \times 4 \times 4$), 4 intransitives (1×4) are possible. Of the 29, there were 18 ditransitives, 9 transitives, and 2 intransitives. A similar ratio was used for the testing corpora.

An additional set of novel sentences was created for the testing phase. For each language there were 20 grammatical items and 20 ungrammatical items. In order to ensure that the FWO languages were not more “regular” than the SWO languages in terms of distributional information, and therefore potentially easier to learn, we controlled our stimuli for five different kinds of fragment information. In the examples below, letter strings correspond to sentences and individual letters correspond to words. For each of the five controlled fragment variables, we computed two differences—between SWO-grammatical items and FWO-grammatical, and between SWO-

⁷ ByteCool’s TextSound program can be found at <http://www.bytecool.com/dltextsnd.htm>

ungrammatical from FWO-ungrammatical. The grammatical and ungrammatical values were then compared using a two-sample t-test.

1) *Associative chunk strength* is measured as the sum of the frequency of occurrence in the training items of each of the fragments in a test item, weighted by the number of fragments in that item (Knowlton & Squire, 1994). For example, the associative chunk strength of the item ZVX would be calculated as the sum of the frequencies of the fragments ZV, VX and ZVX divided by 3. Two-tailed t-tests indicated that there were no differences between the languages in associative chunk strength for the grammatical ungrammatical items ($t=0$).

2) *Anchor strength* is measured as the relative frequency of initial and final fragments in similar anchor positions in the training items (Knowlton & Squire, 1994). For instance, the anchor strength of the item QXMSXV is calculated as the sum of the frequencies of the fragments QX and QXM in initial positions in the training items and of the fragments XV and SXV in final positions in the training items. Again, there were no differences between the two languages in the anchor strength of the grammatical and the ungrammatical items ($|t| < 1$ $p > 0.4$).

3) *Novelty* is measured as the number of fragments that did not appear in any training item (Redington & Chater, 1996); e.g., if the fragments XVS and VS from the item QXVSZM never occurred in a training item, then the test item would receive a novelty score of 2. There were no differences between the grammatical ungrammatical items ($t=0$).

4) *Novel fragment position* is measured as the number of fragments that occur in novel absolute positions where they did not occur in any training item (Johnstone & Shanks, 1999); e.g., if the fragment VQZ from the item QZVQZV never occurred in this absolute position in any of the training items then this item would be assigned a novel fragment position score of 1. There were no differences between the novel fragment scores for the grammatical and ungrammatical items ($|t| < 1$ $p > 0.6$).

5) *Global similarity* is measured as the number of letters that a test item is different from the nearest training item (Vokey & Brooks, 1992); e.g., if the test item QZM has QZV as its closest training item then it would be assigned a global similarity score of 1. There were no differences between the two languages for the grammatical ($t=0$) and ungrammatical ($|t| < 1$ $p > 0.4$) items.

Ungrammatical testing items were created by replacing the verb in the grammatical items so as to violate the verb-argument structure. For instance, *paazi* represented the intransitive verb and so would only occur in two-word sentences. An ungrammatical sentence might be *Paazi zeka tuzee*, violating the normal usage of the word.

Results and Discussion

Overall, subjects showed testing performance consistent with our hypothesis, with some qualifications. Two-way ANOVAs did not reveal significant main effects of word order ($F(3,24) = 1.26$ $p > 0.2$). There was an almost significant effect of case ($F(3,24) = 3.99$ $p < 0.06$) with the two-case conditions being more difficult than the no-case conditions. There was no significant interaction between the two variables ($F(3,24) = 1.22$ $p > 0.2$). Regression analysis revealed that the testing data were not significantly affected by the subjects' age, sex, or number of languages which they spoke. The finding of a tremendous variance in training performance (range=43.1%-99.98% correct; SD=16.55%) was somewhat surprising as we expected the training task of repeating 2-4 word sentences to be quite easy. We were also surprised by the strong correlation between training and testing performance (Pearson $r=0.61$; $F(1,27)=15.29$ $p < 0.001$), considering that training and testing seemed to rely on different cognitive processes. The former relied on short-term memory, the latter on the ability to figure out rules from examples. Questionnaire responses revealed that the subjects with the best training and testing performance were the only ones who could

verbalize the “rules” used (i.e., “*lami* occurred only in 4-word sentences, *paazi* only in two-word”, etc.).

Even though there was no main effect of word order or case, post-hoc t-tests revealed that only two conditions were significantly above chance level (Fig. 4): SWO-no case (mean=30.14; $t(6)=3.6$ $p<0.02$) and FWO-Case (mean=23; $t(6)=3.15$; $p<0.02$). Testing performance in the other two conditions was not significantly different from chance: SWO-Case (mean=22.57; $t(6)=0.95$ $p>0.37$) and FWO-No case (mean=24.57; $t(6)=2.09$ $p>0.08$). Although we did not have an explicit testing-only condition, pilot data indicated that subjects completing only the testing condition scored at chance levels.

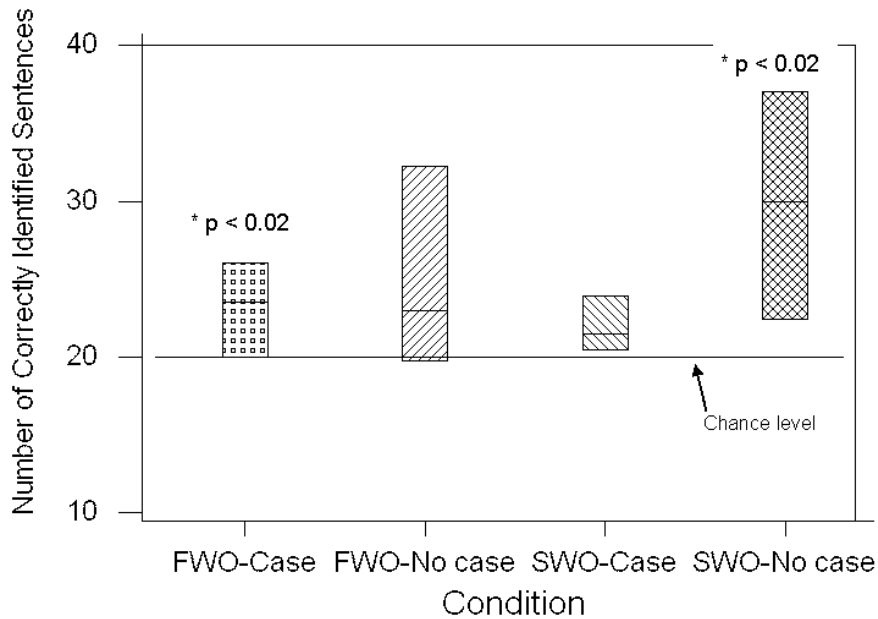


Figure 4: Learning performance for the four grammars. The boxes for each condition show the median along with 1st and 3rd quartiles.

This latter pattern of results is consistent with our hypothesis that rare languages should be more difficult to learn. One qualification we need to make concerns the SWO-Case condition which turned out to be more difficult for the subjects than any other while being the easiest for the networks to acquire. Few languages use both case and word order presumably because when one suffices there’s no reason to develop

the other. Nevertheless, languages with case can support flexible word orders (that is, alternatively-ordered sentences are understandable in case-languages) and so it is likely that such word orders will evolve turning these languages into FWO case grammars. There is therefore no reason to believe that SWO-Case languages should be difficult and we need to explain why our subjects had greater trouble with this condition.

Not only did subjects completing the SWO-Case condition have lower testing scores, but training data revealed that case was a highly significant factor ($F(3,24)=6.74$ $p<0.02$) with subjects making more mistakes on case conditions (mean=38.83% incorrect) as compared to no case conditions (24.08% incorrect). On the surface this may simply indicate that subjects in case conditions had more opportunities to make mistakes, being penalized for incorrect case markings. Analysis of training performance, however, reveals that repetition mistakes in these conditions largely came from the same sources as in the no-case conditions, namely omitting words and confusing words with others.

To explain these results, we need to recognize that case markings added extra complexity to the language by creating more word variants, increasing the processing demands on the subjects. Simple recurrent networks have limited but “perfect” memory and are able to quickly determine that case is a good cue to grammatical category, and requires attention. Human subjects, on the other hand, did not seem to come to this conclusion by the end of training. Although several commented on the order of the words in the post-study questionnaire, not one noticed the pattern of suffixes in the case conditions. We conclude that the most likely reason for the decreased training performance in the case conditions is that cases increased “cost of processing” without offering a clear benefit in acquiring the syntax of the artificial language (see experiment revisions below).

We have yet to explain the strong positive correlation between training and testing performance noted above. Could it be that there is a causal relationship between the two? That is, could better repetition accuracy in training lead to greater accuracy in recognizing grammatical and ungrammatical items as such? Alternatively, is there a mediating variable that can account for individual differences in training and

testing performance? A likely candidate for such a variable is short-term or working memory. But although subjects with better short-term memory are expected to do better on the training component, it is not clear why these same subjects should do better in testing since short-term memory does not help one deal with novel sentences of the type encountered during testing. It is also not obvious why only the subjects with excellent training performance were able to verbalize the rules used.

The most interesting possibility is that successful repetition of training stimuli improves learning. While rote memorizing of training stimuli by itself would not help on the testing (since all the testing sentences are novel), perhaps successful repetition of the training stimuli allows the subjects to hold the sentence in memory for a longer time, thereby facilitating analysis. At the same time, unsuccessful repetition not only does not help in learning the language patterns, but actually degrades learning through proactive interference. A similar but considerably weaker correlation between training and testing performance in artificial grammar experiments was reported by Chris Conway (pers. comm.). In his series of experiments, however, the training task consisted of match/mismatch decisions which forced subjects to pay close attention to the stimuli, but not reproduce them. The presence of a correlation in his studies suggests that perhaps a more likely mediating variable is *representational ability*. Better representation of sequences of shapes, tones, tactile patterns (Conway & Christiansen, 2002) or nonsense words (present study) implies greater opportunity for mental analysis, which in turn leads to better learning. More demanding training tasks such as prompting subjects to actually reproduce the stimuli, may better tap into the subjects' representational capacity, producing a stronger correlation. A confirmation of these results by more robust procedures would open up an exciting avenue of research into the effects of mental representation on statistical learning.

Limitations

Although the results provide preliminary support for our hypothesis, the present experiment has a major shortcoming—the subjects were not required to map the words in the training sentences onto any kind of categories. In a sense, this is equivalent to a

child attempting to acquire her first language listening to a radio broadcast. Although the broadcast may contain all the necessary linguistic structure, that structure is not grounded in a real-world representation. Without a grounding in “reality” a child that does not already have the correct mappings can never get off the symbolic merry-go-round (Harnad, 1990).

It is clear that in order to acquire the “who did what to whom” syntactic relations, some word-to-category mapping needs to occur. To address this issue we plan to revise the experiment as follows.

The training procedure will be similar to the original, except that the auditory sentences will be presented concurrently with visual scenes. The precise nature of the scenes is still to be determined. Figure 5 shows a possible instantiation of this procedure. In this instance, the subject hears the caseless VSO transitive sentence *Tila voopoo soogee*. At the same time, she is presented with a display showing one shape (*voopoo*) acting on the other (*soogee*). In the case of ditransitive sentences, subjects will see an interaction between three shapes, and during intransitive sentences, one shape with an arrow indicating that an action is happening (e.g., John walks). The visuals will be identical in all conditions reflecting the idea that children use similar non-linguistic cues to ground syntactic relationships regardless of the type of language they are learning.

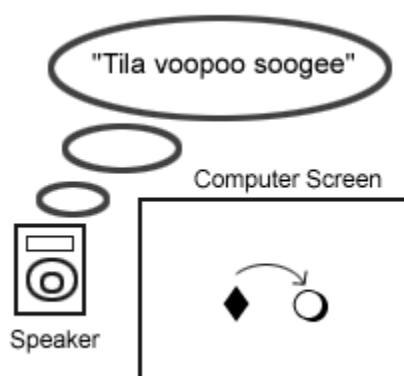


Figure 5: A possible setup for a revised experiment

The auditory and visual stimuli will be presented in pairs—half of the time the two pairs will be identical (match), and half the time they will be different (mismatch). Half of the mismatches will consist of visual mismatches—the first visual will be different from

the second. The remaining half of the time, the auditory portion will not match—the first sentence will be different from the second. After each pair, subjects will be prompted to judge whether the two stimuli were identical or not.

Testing

Subjects will be presented with new sentences and asked to select the appropriate visual from four alternatives. The alternatives will be generated in the following way: one will be correct, one will violate the syntactic relationship by reversing the direction of the arrow. The remaining two will violate the word-to-object mapping as presented in the training part (e.g., *soogee* was always associated with a diamond, *voopoo* with a triangle, and so on). We hypothesize that subjects completing the FWO-Case and SWO-No case conditions will score significantly above chance, while subjects completing the FWO-No case condition will score at chance.

Part III: Language as a Replicator

Throughout this thesis I have assumed that languages change over time. This claim is uncontroversial. In exploring the evolution of language, some of the most fundamental questions are what drives this linguistic change, and why some changes are preferred over others. The empirical data presented so far suggest that learnability constraints play an important role. Consistency seems to be a major factor: with languages employing consistent cues being acquired easier than less consistent ones by simple recurrent networks, human subjects, and children acquiring natural languages (Slobin & Bever, 1982). I have presented my arguments in an evolutionary framework by approaching language as an organism (e.g., Christiansen, 1994). Just as genes that increase the reproduction of their host further their own spread, so syntactic structures

that make languages more learnable persist, while those that are difficult to acquire die out or never come into existence. In the remainder of this discussion I will extend the evolutionary framework to provide an answer to the basic question of the driving force behind linguistic change.

Like genetic complexes, languages mutate through imperfect copying. The imperial efforts of Great Britain spread English throughout the world. In the years that followed, imperfect copying and contact with other languages (an approximate analog of genetic crossovers that occur in sexual mating) has resulted in pronunciation and word differences that produced American and Australian English as well as numerous English creole languages. On a large geographic and temporal timescale such changes eventually lead to the formation of entirely new languages, and often widely different languages (in the case of Indo-European languages: Russian, Hindi, French, and Armenian). Just as genetic change involves random components like mutation and genetic drift, so linguistic change involves analogous components of mutation through imperfect copying, and linguistic drift characteristic of populations separated by space and time.

In addition to more or less random factors, such as mutation and drift, change also arises from language contact. Just as offspring imperfectly resemble their parents, so languages imperfectly resemble their linguistic parents. It has been suggested that language contact more often than not results in morphological simplifications even when the participants already share similar category systems (e.g., Sankoff, 2001, p.16). Such changes can be seen in the evolution of the English language from Old to Modern where a full case system was gradually eliminated through contact with the Normans and the Norman French language (Morris et al., in press).

The linguistic changes presented so far, for instance those that produced Modern English from Old English, operated in a fully linguistic environment. A modern human can acquire Old English, and a native speaker of Old English would have no problems acquiring modern languages. Supporters of an innate universal grammar would explain such acquisition by stating that both languages can be acquired because they are compatible with universal grammar. This reasoning breaks down as we go farther back in time. If UG is specific to language, how could it have existed before humans were

capable of language? If UG formed gradually, as claimed by Pinker and Bloom (1990), then there must have been a time during which humans were capable of acquiring a simpler language but not be able to acquire modern languages, since the latter make use of UG principles not yet present in the mind of those early language-users.

Such an innately prescribed Universal Grammar presents an obvious problem to language evolution. To illustrate: suppose we have a proto-linguistic population in which a genetic mutant is born who is capable of, say, verbalizing the subjunctive tense (i.e., a device to delineate hypothetical events). Because the mutant's ability is not shared by other members of the population, it is not clear how such a mutation could ever spread in a population.⁸ The linguistic status-quo (in the language of theoretical biology) is an uninvadable evolutionary stable strategy. The success of a language in a population depends on mutual comprehensibility. A linguistic mutant will at best be ignored; at worst, be killed for being different. A mutation, no matter how seemingly useful, cannot invade a population if its benefit cannot be made apparent to others.⁹ Even more problematically—*usefulness* is irrelevant when we speak of Darwinian fitness which is, of course, measured in the number of surviving, reproducing offspring. Even if we accept that mutations may be small and produce linguistic changes comprehensible to other, non-mutated individuals, it is not at all clear why a mutation leading to the subjunctive case would ever confer greater Darwinian fitness on its bearer.¹⁰

⁸ Some have argued that because early humans presumably lived in kin groups, the mutation would be shared with some other individuals due to the high relatedness of the group. While this could be true, it is pure speculation unable to stand to even mild inquiry. For instance, what happens if the mutation occurs in a female who leaves the group to mate with a member of another group? More fundamentally, the group has to presumably communicate to individuals from other groups. If each group evolves its own UG, such communication would be impossible.

⁹ One can imagine a mutation in language *perception* whose selective advantage is relatively irrespective of the population.

¹⁰ Claims have been made that tribal leaders have both high reproductive success and are typically gifted orators. This implies that females select males with better language skills, thereby potentially solving the question of the selective advantage of something like the subjunctive—at least as long it could be understood by others in the group. This type of sexual selection would imply that males should have considerably better language skills than females due to stronger selection pressures on the former. This is simply not the case. Men and women have comparable language abilities and studies that do show differences tend to point in the female's favor.

Language as a Cultural Invention

The problems described in the section above can be solved by approaching language as a cultural invention by “language-ready” individuals.¹¹ Language is not based in some mysterious providential system of universal grammar, but is mediated by more general sequence-learning devices. Let’s see how the problems can be solved one-by-one. The same “discovery” of the subjunctive now does not necessitate any genetic mutation. You as well as your conversational partners possess, to use Hombert & Coupé’s term (2002), the *cognitive potential* to imagine the hypothetical. Your discovery was a way to verbalize it. This discovery can spread a great deal quicker than can a mutation through genetic means. One of the greatest shortcomings of genetic approaches to language evolution is the short timescale in which such changes are presumed to have occurred. Language is often assumed to have arisen 40-100 thousand years ago. Even the more conservative figure leaves little time for a full UG to emerge through genetic means. One needs to only look at cultural change in the past two centuries to appreciate the relative speed and power of cultural evolution.

Cultural evolution also frees us from the suffocating requirement of the bearer of the mutation to have more offspring for the mutation to spread. All that is required is that the change somehow catch on in the population. Why should some changes catch on more than others? The first consideration is one with which this thesis has dealt with extensively—learnability. Language features which are more easily learnable are likely to catch on more than those that are more difficult to learn. However, this is insufficient. It’s probably easier to learn to transpose the last two syllables of every word than it is to acquire a subjunctive case. What makes one of these more likely to persist than the other? To answer this question, we need to take the meme’s eye view of language.¹²

¹¹ The term “language-ready,” to denote humans capable of learning language, was used by Arbib, 2002.

¹² According to Dawkins, “memes” are “tunes, ideas, catch-phrases, clothes fashions, ways of making pots or building arches. Just as genes propagate themselves in the gene pool by leaping from body to body via sperms or eggs, so memes propagate themselves in the meme pool by leaping from brain to brain via a process which, in the broad sense, can be called imitation” (Dawkins, 1989, p. 192). In short, memes are cultural self-replicators.

The Selfish Meme Approach

Imagine a meme for some linguistic feature. For the sake of consistency, a meme for the subjunctive. Suppose also that it is also fairly easy to acquire, at least as easy as the other language memes circulating in the population. I would like to make the claim that this meme will spread if it allows or encourages its bearer to express their cognitive state better than before. The intuition for this claim comes from Blackmore (1997) and is quite simple. Imagine a population in which some individuals talk more than others. Focusing on just the memes that are passed on through language, whose will be passed on? The answer is obvious: all other factors being equal, individuals who talk more will pass on more memes, *including* the so-called blabbermouth meme—that is, the *idea* of being a blabbermouth. Clearly, the blabbermouth meme has more opportunity for being spread than a meme that encourages its bearer to stay silent. Why should the meme for the subjunctive have a higher probability of spreading than one dictating to, say, transpose the last two syllables of every word? Simply because the subjunctive meme’s “infestation” of a mind will increase its reproduction by enabling its bearer to express their cognitive state to others in an enhanced way. This may seem circular at first, but on second glance it is the very bedrock of evolutionary theory. If a gene happens to increase an organism’s reproductive fitness, it increases its chance of being passed on (along with other genes contained in the organism). If a meme happens to increase an organism’s ability to pass memes to others, it increases its chance of being passed on (along with other memes contained in the host). Being able to think of hypothetical situations but not verbalize them linguistically implies that once a way to verbalize them becomes available, it will spread. More generally:

*Passing the ability (meme L) to pass concepts to other individuals necessarily increases the prevalence of meme L.*¹³

Given the cultural evolution patterns described above, one question still remains. How does a cognitive system supporting language ever get off the ground? Just as a phone is quite useless if one has no one to call, so language would seem to be of little value to early users. The next and final section presents a possible solution to this problem by expanding on Dennett's (1991) framework of cognitive autostimulation.

Part IV: Language as a cognitive self-stimulator

Many animals have learned to take advantage of the alerting effect that broadband sounds such as screams have on the nervous system (Owren & Rendall, 2001). It is unclear which came first. Is the noxiousness of these sounds an evolutionary adaptation in its own right, or was sound production first—an involuntary consequence of being in a dire situation? The answer is likely to involve co-evolution of both the production and perception systems, each bootstrapping the other. The usefulness of being able to warn others of a present danger, and being able to take advantage of such warnings, is clearly adaptive (in the strictest Darwinian sense). In this section, I will present a view of language as an extension of this system, with a twist. Just as alert calls capitalize on the involuntary response of the receiver, so language capitalizes on the *cognitive* properties of other minds.

In order for (proto) linguistic utterances to have an effect on the cognitive processes of others, an obvious precondition needs to be met. The words must be understood. Unlike an alarm call that can be “hardwired” (thereby preventing or hindering the learning of new calls); the arbitrary nature of words in each language necessitates that the words be learned.

¹³ Strictly speaking, a meme is not an ability, but an *idea*. Ideas, however, can be equivalent to abilities when they are learned by individuals capable of executing them. More specifically, the meme (idea) for the subjunctive is equivalent to the ability to use it as long as it learned by an individual capable of executing this linguistic task.

The Language Spark

Given a society of early humans with most neural abilities for language, what is the spark that started the “Good Trick” (Dennett, 1991) that is language? Dennett explores a fascinating possibility: language as a cognitive self-stimulator. In other words, language helps us think just by verbalizing, or even thinking linguistically to oneself. This has several implications. The first is that “words” were not social, but rather personal constructions. Suppose first words reflected the mental states of their producers. So, the early *Homo sapien* Ug, frustrated by his spear missing the target, screams *grrr*.¹⁴ In doing so, an association is made between his state of frustration and the (arbitrary) sound that comes out. Og, who overheard Ug exclaiming *grrr*, also makes the connection between Ug’s frustration and his utterance.¹⁵ On his next hunting trip Og’s spear may miss and in *his* frustration, influenced by Ug’s exclamation, he may produce a similar one (*garr*), or perhaps one that is completely different (*damn*). However, it is my claim that the next time Ug misses, he will be much more likely to again say *grrr* having the association between the mental state and the action already in his mind. This line of reasoning is inspired by associative Hebbian learning—the idea that repeated firing of one neuron onto another strengthens the connection between the two neurons. According to this principle, the strengthening of the association between *grrr* and the mental state of frustration in this case, is implicit in the fundamental rules of learning.¹⁶

¹⁴ Or for fans of gestural origins of language, Ug drops his hand.

¹⁵ I believe it is safe to assume that prelinguistic humans had at least a simplified theory of mind; that is, they could hypothesize that others have mental states similar to their own (e.g., Premack, 1990b).

¹⁶ Hebbian learning, although simple in principle has many subtle implications. Consider the well-known case of Japanese native speakers learning English and being unable to correctly pronounce /r/ and /l/. Counter to what cursory descriptions of the phenomenon describe, it is not the case that Japanese are unable to produce /r/ and /l/, it is that they use the two phonemes interchangeably and inconsistently, since in Japanese the acoustic difference between them is not phonetic, meaning there are no words in Japanese that differ only in these two sounds (as in *rice* and *lice*). We can use the Hebbian paradigm to gain insight into a possible reason Japanese have so much trouble with learning the correct pronunciation, and why instruction does not seem to help. The reason for the latter is quite startling—instruction can’t help because the process of *categorizing* the phoneme incorrectly upon hearing it reinforces those incorrect pathways by the mere act of categorization (James McClelland, personal communication).

The utterance *grrr* becomes a meaningful signal to Ug, despite its private, subjective nature. What of it? If language is a symbiotic meme, shouldn't it be of some cognitive benefit? Yes, and it is.

Ninio (1990) cites Wittgenstein: "a private language is a contradiction in terms" (p. 746). If the *only* purpose of language is to communicate to *others*, that is, of course, true. But perhaps it isn't.¹⁷ Consider the possibility of using language as a cognitive self-stimulator—a kind of a high-level programming language that can help us organize our mental processes (Dennett, 1991). Dennett quotes Straight: "Cognitive autostimulation is not just some wacky idea a philosopher thought of—it's something we experience all the time" (p. 298).

In itself the idea that language, even private language, helps us think is nothing new (e.g., Bickerton, 2002). Although some individuals describe themselves as thinking in more visual than linguistic terms, few would dispute that language is quite essential to thought. I am not suggesting that natural language is itself the language of thought, an idea to which Fodor replied: "The only thing wrong with this proposal is that it isn't possible to take it seriously" (1975, p. 56). Clearly, we can have thoughts that do not easily translate into words (the texture of brushed aluminum, the sound of a violin, the smell of car exhaust¹⁸). My assumption is only that language *helps* us think, perhaps in a way so fundamental that we take it for granted. Counter to what nativists would like to believe, there is no hard evidence that the "linguistic system" can ever be dissociated from the rest of cognition (Elman, 1999; also see the examples of aphasia, Parkinson's, and family KE above). We find ourselves hard-pressed to imagine what thinking without language is like.

¹⁷ Curiously, Chomsky recently remarked that as far as he is concerned, "the proportion of language used for communication is at the level of statistical noise" (*The 4th International Conference on Evolution of Language*, 2002. Cambridge, MA.). Although his comment is spurious, this section seriously considers the possibility that the early impetus for linguistic organization may have been more for organizing one's thoughts than communicating them to others.

¹⁸ One can argue, however, that by simply referring to these ideas, I transcended their supposed extra-linguistic nature. After all, is my saying "imagine the sound of a violin" any less descriptive than the more concrete "imagine an apple"?

The ideas that language helps us think generally assume a language that is shared by a population. Speaking to others is an exchange of information which “rewires” the thought processes of our brains in potentially constructive ways. But if Language (the idea of language) is to spread through the population, it would help for it to be cognitively useful to its producers even without being understood by others. This is met with obvious criticism: how is talking to oneself useful? We already know everything we are going to say! Dennett helps us clear up this confusion. The flaw with the objection to self-verbalizing comes from assuming a central-executive, an omniscient homunculus that is somehow privy to everything we are about to say. There is no “I” module in the brain (see Dennett, 1991; Minsky, 1986; Kosslyn, 1994). All that exists are brain faculties (brain regions), some of which are well interconnected and share knowledge. Others are less well connected and may be able to benefit from “hearing” verbalizations from these neurally distant regions. In Dennett’s words: “such an act of autostimulation could blaze a valuable new trail between one’s internal components” (p. 196).

Let’s go back to Ug, who now has a pairing between *grrr* and a sense of frustration. He is now able to prime that feeling by simply saying *grrr*. Of what use is it? Perhaps he happens to say it again in the context of practicing spear-throwing and missing. This may remind him of the failure of the hunt and encourage him to practice more.¹⁹ But more importantly, he is now potentially able to prime a mental state in others who have observed (and remembered) him say *grrr* in frustration; That is, he can make others frustrated by uttering a sound which is in itself meaningless! In time, it may become adopted in his hunting party as a thematic marker—describing the situation in which one misses an animal with a spear—a likely scenario in a hunter-gatherer society.

In a series of recent simulations, Kirby and Hurford (2001) have shown how syntax can emerge from more holistic utterances. Given the right parameters, discrete words become assigned to discrete concepts. *Grrr* should at some point stop applying

¹⁹ Because this framework does not rely on natural selection *per se*, we are not under pressure to come up with an explanation for how *grrr*’ing increases one’s reproductive fitness. A scenario in which it provides an immediate (and not necessarily reproductive) advantage is quite sufficient. It is in Dennett’s words, a “Good Trick”.

to the holistic situation being described, and come to mean “I am frustrated” or simply “frustrated.” Discreteness, however, is not a clear-cut concept. The Native American language Nootka has a discrete word *tlih* for “moving pointwise like a docked canoe,” a concept that is not discrete in English. In Kirby and Hurford’s simulations, the emergence of a context-free verb “to eat” is seen as an improvement over a word that is totally different depending on whether John eats or Mary eats. But consider a language like Shawnee, which uses a single word *kwaskwi* to denote a condition of recoil: two forces pushing and repelling from each other. Using this type of syntax, “I push his head back” and “I drop it in water and it floats” result in nearly identical sentences with the difference being a focus on the head as opposed to the water. In such a language “eat” may be a different word depending on who is doing the eating and in what circumstances. In contrast, there may be a single word to describe anything that is forked, or anything that “moves downwards.” For instance, in Apache “it is a dripping spring” becomes an agglutination of *ga-* (whiteness or clearness), *no-* (the downwards motion), and *-tó* (water or spring) (Whorf, 1952, p. 21).²⁰ Both systems are presumably efficient (otherwise they would not exist so widely), but the latter has a quality of priming that the former does not.

Coming back to our early humans, consider how many concepts can be formed with *grrr* as a root originally meaning the frustration arising from a missed spear-throw during a hunt. *Grr-ui-ra* may mean, “The frustration stemming from your friend forgetting to meet you at the waterhole. *Grr-ad-ni-sak* may come to mean, “the way a woman’s eyes are shaped when she is annoyed with her husband.” What do they have in common? Their root—*grr*—primes the theme of frustration. In direct contrast, English sentences about boat-objects are likely to share a common phonology (most would make direct references to water, boats, etc.). In Nootka, however, the “words” *tlih* and *lash*, instead of priming the taxonomic category of boats, prime thematic categories of a type of movement (characteristic of boats, but also as applied to unrelated objects), and of people that pilot boats, respectively.

²⁰ To be fair, English also has compound terms like “waterfall,” but they account for a small minority of the language.

In attempting to explain how humans come to categorize their world into hierarchical systems, Markman and Hutchinson (1984) have described how the linking of words with objects shifts categorization from theme based to taxonomy based. A young child will tend to group a dog and a bone together in a thematic relationship—dogs eat bones. However, when older children are shown a dog and told it's called a “dax” and then asked to find another dax, they will put a dog and a cat together—a grouping based on taxonomy. This taxonomic shift, often taken for granted in literature, never needs to happen because the primary purpose of many Native American words I have mentioned is not to refer to objects, but rather to prime themes.²¹

In Hopi, a single noun is shared by all flying objects with the exception of birds (Whorf, 1952, p. 6). So while in English “mosquito” may psychologically prime “flight,” there is nothing in the phonology of one to imply the other. Similarly, the onset of “captain” primes “cap” (e.g., see Max Planck Institute Annual Report), but the words are semantically unrelated—the phonological priming here is a “mistake” because it does not facilitate understanding.²² In a language like Shawnee or Nootka such priming is semantically correct! In such a language if *cap* referred to the quality of sitting on one's head like a hat, then words beginning or even containing *cap*²³ would refer to a similar quality. This agglutinating property can cause inconceivable results. For instance, in Shawnee the “word” *l'θawa* refers to a “forking outline” (Whorf, 1952, p. 18). Because an extra toe branches out from a foot like twigs from a tree, the sentences “I pull the branch aside” and “I have an extra toe on my foot” are nearly identical. It seems clear that “primary words” in such a language are not object names. They do not even *refer* as such. They are primes—activating thematic categories.

To review: Given that the neural substrates for language are established by some non-linguistic-based process of natural selection, language may first arise from cognitive auto-stimulation. Individuals may verbalize their mental state, and later use it to bring forth the mental state in themselves in useful ways. This is not as absurd as it

²¹ Interestingly, Premack (1990a) cites a study by Smiley and Brown that showed adults of 60 and older to revert back to thematic sorting. This suggests that perhaps it is not language as such that causes the taxonomic shift in young children.

²² The word *captain* incidentally derives from the Latin root *capit-* meaning *head*. (dictionary.com)

may sound: consider how telling a joke can make you yourself laugh. Presumably you already knew the punch line, and yet somehow, the telling of the joke, or thinking back to a funny situation, can change your affective state. The utterances at first have private meanings, but in time, they are picked up by observers through ostensive and non-ostensive methods (e.g., Tomasello & Akhtar, 1995)—the thematic meanings begin to be shared between people. In time, the syntax is embellished with more precise prefixes and suffixes (Newmeyer, 2000). With the increased complexity that comes with dealing with many “words,” grammatical conventions are reinforced (see Kirby & Hurford, 2001 for discussion), and constructions become increasingly stylized and streamlined through cultural exchange. The language meme is on the loose, and free to make us human.

²³ The thematic root of such word-sentence tends to be at the beginning (Whorf, 1952).

References:

- Aitchison, J. (1996). *The Seeds of Speech: Language Origin and Evolution*. Cambridge: Cambridge University Press.
- Arbib, M.A. (2002). Language-readiness and the evolving mirror system. To appear in M.H. Christiansen & S. Kirby (Eds.), *Language evolution: States of the art*. Oxford University Press.
- Bellugi, U. (1983). Language Structure and Language Breakdown in American Sign Language. In M. Studdert-Kennedy Ed., *Psychobiology of Language*. MIT Press, Cambridge MA
- Blackmore, S. (1997). The Power of the Meme Meme. *Skeptic* 5(2): 43-49.
- British National Corpus*. Located at: www.hcu.ox.ac.uk/BNC
- Calvin, W.H., & Bickerton, D. (2000). *Lingua ex Machina: Reconciling Darwin and Chomsky with the human brain*. MIT Press, Cambridge MA
- Cassidy, K.W. & Kelly, M.H. (1991). Phonological information for grammatical category assignments. *Journal of Memory and Language*, 30: 348-369.
- Cavalli-Sforza, L.L. & Feldman M.W. (1981). *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton University Press.
- Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.
- Chomsky, N. (1988). *Language and Problems of Knowledge: The Managua Lectures*, Cambridge: MIT Press, Cambridge MA.
- Christiansen, M. H. (1994). *Infinite languages, finite minds: Connectionism, learning and linguistic structure*. Unpublished doctoral dissertation, Centre for Cognitive Science, University of Edinburgh, U. K.
- Christiansen, M.H. & Dale, R.A.C. (2001). Integrating distributional, prosodic and phonological information in a connectionist model of language acquisition. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 220-225). Mahwah, NJ: Lawrence Erlbaum.
- Christiansen, M.H. & Devlin, J.T. (1997). Recursive inconsistencies are hard to learn: A connectionist perspective on universal word order correlations. In *The Proceedings of the 22nd Annual Conference of the Cognitive Science Society*. (pp. 113-118). Mahwah, NJ: Erlbaum
- Conway, C.M. & Christiansen, M.H. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences*, 5, 529-546.
- Conway, C.M., & Christiansen, M.H. (in press). Sequential learning by touch, vision, and audition. *Twenty-fourth Annual Meeting of the Cognitive Science Society*.
- Corballis, M.C. (2002a). From hand to mouth: The gestural origins of language. To appear in M.H. Christiansen & S. Kirby (Eds.), *Language evolution: States of the art*. Oxford University Press.
- Corballis, M.C. (2002b). *How gestures became vocal*. Paper presented at the Fourth International Conference on the Evolution of Language. Cambridge, MA.
- Cottrell, G.W. (1989). *A Connectionist Approach to Word Sense Disambiguation*. Research Notes in Artificial Intelligence. Morgan Kaufmann, San Mateo.
- Dawkins, R. (1989) *The Selfish Gene*, 2nd edn. Oxford: Oxford University Press.

- Deacon, T. (1997). *The Symbolic Species: The co-evolution of language and the human brain*. Penguin Press.
- Dennett, D. (1991). *Consciousness Explained*. Boston: Little, Brown
- Dunbar, R. (2002). The social origins of language. To appear in M.H. Christiansen & S. Kirby (Eds.), *Language evolution: States of the art*. Oxford University Press.
- Ellefson, M.R. & Christiansen, M.H. (2000). Subjacency constraints without universal grammar: Evidence from artificial language learning and connectionist modeling. In *The Proceedings of the 22nd Annual Conference of the Cognitive Science Society*. (pp. 645-650). Mahwah, NJ: Erlbaum
- Elman, J. L. Origins of language: A conspiracy theory. In B. MacWhinney (Ed.). *The emergence of language*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Elman, J.L. (1990) Finding structures in time. *Cognitive Science*, 14:179-211.
- Fodor, J.A. (1975). *The Language of Thought*. New York: Crowell.
- Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, 3, 419-429.
- Gopnik & Crago (1991), Familial aggregations of a developmental language disorder. *Cognition* 39, 1-50
- Greenberg, J. H. (1963) Some universals of grammar with particular reference to the order of meaningful elements. In: *Universals of Language*. Greenberg, J.H. (ed.). MIT Press, Cambridge MA
- Hare, M. & Elman, J.L. (1995). Learning and morphological change. *Cognition* 56: 61-98
- Harnad, S. (1990) The Symbol Grounding Problem. *Physica D* 42:335-346.
- Hombert, J.M., & Coupé, C. (2002). Language polygenesis: What type of evidence?. Paper presented at the *Fourth International Conference on Evolution of Language*. Cambridge, MA.
- Jacennik, B. & Dryer, M.S. (1992). Verb-Subject Order in Polish. In *Pragmatics of Word Order Flexibility*, ed. Payne, D. L. Amsterdam: John Benjamins.
- Johnstone, T., & Shanks, D. R. (1999). Two mechanisms in implicit artificial grammar learning? Comment on Meulemans and Van der Linden (1997). *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 25(2), 524–531.
- Kay, R.F., Cartmill, M. & Balow, M. (1998). The Hypoglossal Canal and the Origin of Human Vocal Behavior. *Proceedings of the National Academy of Sciences* 95(9):5417-5419.
- Kirby, S. (1998). Language evolution without natural selection: From vocabulary to syntax in a population of learners. Edinburgh Occasional Paper in Linguistics, EOPL-98-1.
- Kirby, S. & Hurford, J.R. (2001). The emergence of linguistic structure: An overview of the iterated learning model. In Cangelosi A. & Parisi D. (Eds), *Simulating the evolution of language*. London: Springer Verlag.
- Knight, C. Studdert-Kennedy, M. & Hurford, J. Eds., (2000). *The Evolutionary Emergence of Language*. Cambridge University Press.
- Knowlton, B. J, & Squire, L. R. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 20(1), 79-91.

- Koster, J. (1999) The word orders of English and Dutch: Collective vs. individual checking. In: *Groninger Arbeiten zur germanistischen Linguistik*, Abraham, W. (ed.), University of Groningen, Groningen. 1-42.
- Kosslyn, S.M. (1994). *Image and Brain: The Resolution of the Imagery Debate*. MIT Press: Cambridge, MA
- Lai, C.S.L., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F., Moncao, A.P. (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*. 413: 519-523
- Lieberman, P. (2000). *Human Language and our Reptilian Brain*. Harvard University Press. Cambridge, MA.
- Lieberman, P. (2002). A biological view of the evolution of language. To appear in M.H. Christiansen & S. Kirby (Eds.), *Language evolution: States of the art*. Oxford University Press.
- Markman, E. & Hutchinson, J.E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic vs. thematic relations. *Cognitive Psychology*, **16** 1-27.
- McDonald J.L. & Plauché, M. (1995). Single and correlated cues in an artificial language learning paradigm. *Language and Speech* 38(3), 223-236
- Minsky (1986). *The Society of Mind*. New York: Simon and Schuster
- Morris W.C., Cottrell, G.W., & Elman, J.L. (in press). A connectionist simulation of the empirical acquisition of grammatical relations. In *Hybrid Neural Symbolic Integration*. Wermter, S. & Sun, R. (Eds.) Springer Verlag
- Neeleman, A. (1994) *Complex Predicates*. PhD Dissertation. Utrecht University, Utrecht.
- Newmeyer, F.J. (2000). On the Reconstruction of 'Proto-World' Word Order. In *The Evolutionary Emergence of Language*. Knight, C., Hurford J.R., Studdert-Kennedy, M. Eds. Cambridge University Press.
- Ninio, A. (1990). The genome might as well store the entire language in the environment. *Behavioral and Brain Sciences* 13(4): 746-747. (Commentary on Pinker and Bloom: same issue)
- Owren, M.J. & Rendall, D. (2001). Sound on the rebound: Returning form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evolutionary Anthropology*, 10, 58-71.
- Payne, D. L.(1992). Verb-Subject Order in Polish. In *Pragmatics of Word Order Flexibility*, Amsterdam: John Benjamins.
- Pinker, S. (1994). *The language instinct: How the mind creates language*. New York: William Morrow.
- Pinker, S. (1995). Language Acquisition In *An invitation to cognitive science. Vol 1: Language* Gleitman, L.R. & Lieberman, M. (Eds.) MIT Press, Cambridge MA.
- Pinker, S. & Bloom, P (1990). Natural language and natural selection. *Behavioral and Brain Sciences* 13(4):707-84.
- Plaut, D. C. & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, 10, 377-500.
- Plunkett, K. & Elman, J.L. (1997). *Exercises in Rethinking Innateness: A Handbook for Connectionist Simulations*. MIT Press.

- Premack, D. (1986). *Gavagai! or the Future History of the Animal Language Controversy*. Cambridge, MA: MIT Press.
- Premack, D. (1990). On the coevolution of language and social competence. *Behavioral and Brain Sciences* 13(4): 754-755. (Commentary on Pinker and Bloom: same issue)
- Quine, W.V.O. (1960). *Word and Object*. MIT Press, Cambridge MA.
- Ratner, B.N., Gleason, B.J., Narasimhan, B. (1998). Introduction to psycholinguistics: What do language users know? In J. Berko Gleason & N. Bernstein Ratner (Eds.), *Psycholinguistics* (2nd ed., pp. 1-40). Orlando, FL: Harcourt Brace.
- Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: A reevaluation. *Journal of Experimental Psychology: General*. 125(2), 123-138.
- Rumelhart, D. E. and McClelland, J. L. (1986). On learning the past tenses of English verbs. In *Parallel Distributed Processing; Volume 2: Psychological and Biological Models*. pp. 216-271. MIT Press, Cambridge, MA.
- Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926-1928.
- Sankoff, G. (2001). Linguistic Outcomes of Language Contact. In *Handbook for Sociolinguists*. Chambers, Trudgill, Schilling-Estes (Eds). Basil Blackwell.
- Slobin, D. I. (1966). The acquisition of Russian as a native language. in F. Smith and G. A. Miller (eds.), *The Genesis of Language: A Psycholinguistic Approach*. Cambridge, MA.: MIT Press
- Slobin, D. I., & Bever, T. G. (1982). Children use canonical sentence schemas: A crosslinguistic study of word order and inflections. *Cognition* 12: 229-265
- Tomasello, M. & Akhtar, N. (1995). Two-year-olds use pragmatic cues to differentiate reference to objects and actions. *Cognitive Development*, 10, 201-224.
- Tomasello, M. & Barton, M. (1994). Learning words in non-ostensive contexts. *Developmental Psychology*, 30, 639-650.
- Tomasello, M., Strosberg, R., & Akhtar, N. (1996). Eighteen-month-old children learn words in non-ostensive contexts. *Journal of Child Language*, 23, 157-176.
- Van Everbroeck, E. (1999). Language type frequency and learnability. A connectionist approach. In *Proceedings of the 21st Annual Conference of the Cognitive Science Society* (pp. 755-760). Mahwah, NJ: Erlbaum
- Van Everbroeck, E. (2001). *Language Type Frequency and Learnability from a Connectionist Perspective*. Submitted Manuscript.
- Vargha-Khadem F, Watkins K, Alcock K, Fletcher P, Passingham R. (1995). Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. *Proceedings of The National Academy of Science*. 92(3): 930-933
- Vokey, J. R., & Brooks, L. R. (1992). Salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 18: 328-344.

Whorf, B.L. (1952). *Collected Papers on Metalinguistics*. Washington D.C.: Department of State. Foreign Service Institute.