

**Changing minds changing tools:
From learning theory to language acquisition to language change**

Vsevolod Kapatsinski

University of Oregon

Table of Contents

Acknowledgments (TBW)

Foreword

1. The web in the spider	1
1.1. Associationist learning mechanisms: Hebbian and error-driven learning	2
1.2. Associative symmetry and hippocampal learning	7
1.3. Setting up the problem: A role for cognitive control	8
1.4. Node properties: Attention, associability, accessibility	9
1.5. What is a domain-general learning mechanism anyways?	11
1.6. Why associationism?	14
1.7. When do you have a channel bias?	17
1.8. Conclusion and preview	21
2. From associative learning to language structure	21
2.1. What we can vs. what we want	23
2.2. Probability matching, maximizing, regularization and the explore/exploit dilemma	24
2.3. Exploration/exploitation and S curves	28
2.4. The framework: Associations between categories	29
2.5. What is to be learned	31
2.6. Directionality	38
2.7. Specificity	42
2.8. Conclusion and preview	45
3. What are the nodes? Unitization and configural learning vs. selective attention	48
3.1. What makes units fuse	49
3.2. Part-whole interactions in learning: Hierarchical inference	53
3.3. Occasion-setting and cue-outcome fusion	54
3.4. Mediation: An open field	57
3.5. Incremental processing and loss of configurality	58
3.6. Elemental comprehension	61
3.7. Selective attention to parts, reanalysis and reduction	62
3.8. Learning to attend	65
3.9. Learning to ignore	67
3.10. Conclusion	70
4. Bayes, rationality, and rashionality	72
4.1. Confidence	73
4.2. Confidence and frequency attenuation	74
4.3. Entrenchment and suspicious coincidence	75
4.4. Learned inattention	77
4.5. Exoneration and Holmesian inference	78
4.6. Outcome competition as covert cue competition	83
4.7. Active learning: What do we seek?	84
4.8. Generative vs. discriminative models	86
4.9. Conclusion	91

5. Continuous dimensions and distributional learning	92
5.1. Within-category variability and extrapolation	95
5.2. Distributions and distributional learning	98
5.3. Salience of the unexpected in distributional learning	104
5.4. Conclusion	113
6. Schematic structure, Hebbian learning and semantic change	114
6.1. Semantic extension vs. entrenchment	117
6.2. The experiment	119
6.3. A Hebbian learning model	125
6.4. Conclusion	135
7. Learning paradigmatic structure	138
7.1. Competition in production	142
7.2. Interactions between the three kinds of structure in morphology	144
7.3. Copying: The operation behind creative production	155
7.4. Conclusion	161
8. Learning morphophonology: The interplay of syntagmatic, schematic and paradigmatic structure	162
8.1. Paradigms and schemas: A review	162
8.2. The data	163
8.3. Naïve Discriminative Learner	168
8.4. Model 1. Paradigmatic and schematic structure	168
8.5. Evidence for copy outcomes	175
8.6. Avoidance vs. Hebbian overgeneralization	179
8.7. Paradigmatic perseveration	182
8.8. The need for syntagmatic structure	183
8.9. Model 2: Syntagmatic structure	186
8.10. An alternative: Generative models	
8.11. Lexical cues	188
8.12. A bias against large changes	192
8.13. A comparison to Clamoring for Blends	195
8.14. Limitations and future directions	197
8.15. Conclusion	198
9. Automatization and sound change	200
9.1. Sequence representations: Hierarchies and chains	200
9.2. Automatization: From hierarchy to chain	202
9.3. Speech: A network of chains	203
9.4. Automatization as an explanation for sound change	205
9.5. Keeping reduction in check	
9.6. Automaticity and interruptibility	207
9.7. Automatization and (in)attention	209
9.8. Automatization vs. unitization	211
9.9. Conclusion	212
10. Bringing it all together	213
10.1. Structure	213

10.2.	Processing	218
10.3.	Acquisition	220
10.4.	Change	222
10.5.	Explaining rarities	228
10.6.	The future	233
10.7.	Conclusion	239
11.	References	241

Foreword

By the time you were just a year old, you had learned what sound distinctions matter and which ones do not. From the constant streams of acoustic and visual input, you had extracted a few acoustic forms and linked them to meanings – your first words. The muscles of your tongue, jaw, and larynx (and a few others) had been shaped into producing intricate, precisely co-ordinated patterns that would re-produce these complex patterns of sound, or at least reproduce them closely enough to evoke the corresponding meaning in the mind of your listener. By three, you were quite adept at producing words and sentences you have never heard before, planning and executing a novel sequence of muscle movements to convey a novel meaning. These feats appear miraculous, impossible for mere animals to accomplish. And indeed, these miracles have led many researchers of language acquisition to posit that we are born knowing much about what human languages are like (Universal Grammar), and specialized learning mechanisms, tailored to the acquisition of language, that are not subject to the laws that govern learning in the rest of the biological world (the Language Acquisition Device). The aim of this book is to convince you that this conclusion is – if not wrong – then at least premature. Language acquisition is simply learning. This book is one illustration of how accepting this proposition gets us much closer to explaining why languages are the way they are – the ultimate goal of linguistic theory – than accepting innate knowledge of language universals and the language acquisition device.

Explaining language by explaining language change

All theoretical linguists are primarily interested in explaining why languages are the way they are. **Usage-based** theoretical linguists like myself approach this question in what may seem like a round-about manner, by explaining why languages change in the ways they do. As emphasized by Bybee (2001), languages are highly diverse, yet they change in predictable ways. Linguistic diversity has often frustrated the search for synchronic linguistic universals to enshrine in a Universal Grammar. Yet, it is fully compatible with *diachronic* universals, recurrent pathways of language change. According to usage-based linguistics, these pathways arise from the way that languages are used and learned, from processes that happen in the moments of perception and production.

In phonology, the search for Universal Grammar culminated in a highly influential approach to linguistics called Optimality Theory, first developed in a 1993 manuscript recently published as Prince & Smolensky (2008). Optimality Theory proposed that the grammars of all languages are comprised of the same finite set of constraints, and the only way in which languages differ is in how they rank these constraints. In phonology, where this theory took hold, the constraints were prohibitions against particular structures (markedness constraints) and prohibitions against various sorts of changes (faithfulness constraints). For example, consider words like *prince*. The [ns] sequence at the end of that word is cross-linguistically rare, and indeed is changing into [nts] in English, so that the word comes to sound the same as *prints*. An Optimality Theorist would therefore be tempted to propose a markedness constraint against this offensive structure, say, *NS. In languages that do have [ns], this constraint would be ranked higher than the faithfulness constraints prohibiting various repairs to the sequence. In

other languages, one of these constraints could be ranked lower than *NS resulting in elimination of the structure. How could it be eliminated? One possibility is violating the universal constraint Max, which prohibits deletion. Demoting Max would turn *prince* into *prin* or *priss*. Another possibility is turning [s] into [t], resulting in *print*. What is chosen instead is violating Dep, which bans insertion. Why was that option chosen, and why was the inserted consonant [t] and not any other consonant, or perhaps a vowel that would break up the [ns] cluster, turning *prince* into *priness*. The theory does not answer.

On a diachronic account, this is not a mystery at all (Browman & Goldstein, 1989). Consider how *prince* is pronounced. In order to transition from a voiced nasal stop [n] into a voiceless oral fricative [s], three things need to happen at exactly the same time. The airflow into the nose needs to be stopped by raising the velum. The voicing needs to cease, which is accomplished by pulling the vocal folds apart to prevent them from vibrating. The tongue needs to be slightly lowered to allow for turbulent airflow between the top of the tongue blade and the hard palate. Note that the first two actions can be done rapidly: no matter how far apart the vocal folds go, voicing will cease; no matter how fast you raise the velum, the airflow will cease. In contrast, the last action of lowering the tongue needs to be quite precise and carefully executed: lower the tongue too much, and you will end up with too much airflow and no frication. Think about jerking a door open or slamming it shut vs. opening it just a crack, just enough for your cat to come in: this last action requires quite a bit more time and premeditation. Now consider what happens if slamming the velum shut and jerking the vocal folds open happen too fast, before the tongue is lowered. What would result is a period of time during which there is no airflow from the mouth or nose and the voicing has ceased; in other words, a [t]. No other repair would just as easily result from a simple articulatory mistiming.

In English, almost all fricatives are allowed in the coda. The sole exception is [h], the fricative that lacks any oral articulation, thus we are allowed *kiss* but not **kih*. In some other languages (e.g. Slave), [h] is the *only* consonant allowed in the coda; *kih* but **kiss*. Synchronically, there seems to be no pattern here at all. In contrast, diachronically, we can say that oral fricatives in the coda position tend to lose their oral articulations becoming [h], and that [h] itself is likely to be lost, reducing to zero. Thus, both languages that have [h] in the coda and lack other fricatives, and those that only allow [h] in this position can result from the same diachronic pathway, {f;s;x} > [h] > 0 (Bybee, 2001: 221). If an old [h] is lost before newer fricatives have reduced into [h], we get languages like English. If coda fricatives have reduced to [h] but no further, we get a language like Slave. In this case, as in many others, much stronger generalizations can be made about the dynamic patterns of language change than about the static patterns of language structure.

In some cases, a strong generalization about synchronic systems *can* be made but is mysterious without considering language change. For example, across languages, adjectives and determiners tend to be on the same side of a noun they modify: *the big cat* vs. **the cat big* or **big cat the*. (There are exceptions, like Spanish.) Like many universals (or near-universals) this one has been promoted to a principle of Universal Grammar, a piece of innate knowledge all human infants inexplicably share. Yet, diachronically, there is no mystery here either. Determiners are not coined by speakers de novo. Rather, they gradually develop out of adjectives through the

process of semantic extension / bleaching inherent to re-use of a frequent form (see Chapter 6). As the meanings of adjectives change, their position in the sentence does not: words tend to grammaticalize 'in situ', rather than hopping around the sentence (Givon, 1979). Eventually, we start calling the frequent, highly bleached adjectives determiners and are puzzled by the synchronic similarity between them and 'true' adjectives. The synchronic similarity in behavior is the result of diachronic similarity in function. Adjectives and determiners stay close because they come from the same place. Like the synchronic patterns of phonology, the synchronic patterns of syntax are explicable only by considering where they come from.

Since there are so few true universals shared by all languages, an important advantage of the diachronic approach is that it explains both the prevalent patterns and the typological rarities. A sequence of perfectly natural changes can result in a system that synchronically looks utterly unnatural. For example, in Russian, [k] and [g] become [tʃ] and [ʒ] respectively before the suffix *-ok*, a palatalization process. Synchronically, this is inexplicable: palatalization happens when velars or alveolars assimilate to front vowels or glides; cf. the [d] in *would you*. The back vowel [o] shares little with [tʃ] and [ʒ]. However, diachronically, this is a simple case of assimilation followed by deletion: the suffix used to be *-jok*, and it is this [j] that caused palatalization before disappearing. Indeed, the English [d] and [j] in *would you* and other frequent phrases is entering the same pathway (Bush, 2001). As part of acquiring one's native language, one has to learn many things that do not 'make sense', that lack a synchronic motivation. It is by looking at how languages are reshaped by the processes of language learning and use that we can explain why the patterns we have to learn are the way they are.

On innateness, learning and learning mechanisms

The primary aim of this book is to integrate usage-based linguistic theory with domain-general learning theory; to bring the body of knowledge on how we learn to bear on the question 'why languages change in the ways they do'. However, much of this work is not based on learning *language*. One may therefore wonder whether it is at all useful for gaining insights into language acquisition. Indeed, Chomsky (1959 et seq) has argued forcefully that language acquisition has to be considered on its own terms, as a unique, species-specific ability governed by its own laws and undergirded by a body of domain-specific innate knowledge.

There is no question that language is both specific to the human species and universal within it. However, a species-specific system of behavior need not require innate knowledge of the behavior. Furthermore, it is uncontroversial that all organisms have specialized learning systems in the sense of distinct neural networks localized to specific parts of the brain and innately prepared to form certain kinds of associations rather than others. However, distinct learning systems need not operate according to unique principles. Distinct neural networks connect different representations, and so embody different systems of belief and conditional behavior. However, the same principles – **laws** or **mechanisms of learning** – can describe how the connection strengths constituting the knowledge of a neural network change on the basis of experience. Laws of learning are constrained by two things: what the system needs to learn, and what it can learn given its biological makeup. Since distinct learning systems are built out of the

same ‘stuff’ – neurons – and need to accomplish the same basic task – learn contingencies between stimuli or between stimuli and behaviors, they have little choice but obey the same laws. Throughout this book, we will see how these laws turn out to be helpful for explaining recurrent patterns in language change.

There are many cases of species-specific behaviors that are acquired by some species fairly rapidly and by others very slowly or not at all. For example, consider song acquisition in male cowbirds (*Molothrus ater*), extensively studied by Andrew King, Meredith West and their colleagues. Cowbird songs are not sung by other bird species, and all male cowbirds end up acquiring a distinctively cowbird song. Importantly, cowbirds are not raised by cowbirds: like cuckoos, cowbirds place their eggs into other birds’ nests. Unsurprisingly, this has initially led to claims of innate song knowledge – much like the claims of innate grammatical knowledge in linguistics (e.g. Mayr, 1974). However, it turned out that cowbird songs are shaped by the females, who prefer the kinds of songs adult males in their community sing (West & King, 2008). Most times a male cowbird sings, he sings to a female. The female does not sing, but indicates her attraction to a male by giving a little twitch of the wing – a good song gives her the shivers, so to speak. The closer the male’s song is to the songs of the adult males in the female’s community, the more likely she is to twitch in appreciation (West et al., 2003). In a novice performer, there is always much random variation in performance – a novice does not hit the target on every try. In the young cowbird, this variation produces a range of variants, some closer to the adult song than others. The closer variants are more likely to be reinforced by the females. In this way, the range of song variants shifts ever closer to the adult song, a process called **shaping** (King et al., 2005). Shaping leads the male cowbird to converge on the type of song preferred by the females he is housed with. Indeed, male cowbirds reared with female canaries incorporate canary elements into their songs – and prefer to court unfamiliar canaries over unfamiliar females of their own species, despite the canaries never reciprocating their advances (King et al., 1996). While much could still be innate about this process, innate song knowledge is not required to explain song acquisition in cowbirds.

Like birdsong, speech is a species-specific behavior that develops in all neurotypical members of the species in typical social environments. Also like cowbird song, speech develops in part by imitating *the right* models. When placed in their normal social environment, male cowbirds are shaped to imitate only adult males rather than (continuing to) reproduce songs of the birds that raised them when they were little chicks (West & King, 2008). Similarly, children learn to speak like their peers rather than like their caregivers (e.g. Kerswill, 1996). The parallels are quite striking.

It is uncontroversial that we are prepared to produce speech: all infants babble, both vocally and manually. Like immature birdsong, early babble is highly variable. Whereas adults find it much easier to produce vocal or manual gestures that are part of their native language – to the extent that other sorts of gestures seldom surface even in error – early gestures do not closely resemble the gestures of the ambient language and are far less differentiated. Over time, gestures that approximate ambient language sounds (and especially words!) are reinforced, making the child more likely to attempt them in the future. With repeated attempts, the reinforced gestures become easier and easier to produce, while the unreinforced gestures