Metathesis in English and Hebrew: A Computational Account of Usage-Based Phonology

Paul De Palma
depalma@gonzaga.edu
Department of Computer Science, Gonzaga University
Spokane, WA 99258-0026

Sara Ganzerli
ganzerli@gonzaga.edu
Department of Civil Engineering, Gonzaga University
Spokane, WA 99258-0026

Shannon Overbay
overbay@gonzaga.edu
Department of Mathematics, Gonzaga University
Spokane, WA 99258-0026

George Luger
luger@cs.unm.edu
Department of Computer Science
University of New Mexico
Albuquerque, NM 87131

Kim Glaspey
kglaspey@zagmail.gonzaga.edu
Department of Computer Science, Gonzaga University
Spokane, WA 99258-0026

Abstract

It is now well understood that language use shapes the acoustic delivery of phonological patterns. One common example of this type of language change-under-use is metathesis, which is the reversal of the expected linear ordering of sounds. The gradual transformation of the Spanish word *chipotle* to *chipolte* in the United States is an example of metathetic change. The Genetic Algorithm (GA) is an optimization technique loosely based on the idea of natural selection. This paper shows that the GA can provide a computational model of a usage-based account of examples of metathesis. In the process, it argues that computer models can bring precision to linguistic theory. As an example we create a GA that is able to characterize metathesis in English and then is able to achieve even better results for related expressions in modern Hebrew.

Keywords: Genetic Algorithm; metathesis; computational phonology; emergent

Usage-Based Linguistics

In the first paragraph of her book on usage-based phonology, Joan Bybee says that "language use plays a role in shaping the form and content of sound systems…[It] affects the nature of mental representation and in some cases the actual phonetic shape of words" (Bybee 2001, p. 1). A non-linguist might reasonably reply, "of course, what else besides use and anatomy could shape sound systems?" Professor Bybee could then show us an elegant but deeply counterintuitive body of work, beginning with that of de Saussure in the early 20th century, which argues that language use can be separated from language competence and, crucially, language competence is where the real action is.

While granting the richness of the formalist program in language study, those of us coming from other disciplines might be pleased to learn that beginning in the mid-nineteen seventies, and especially with the wide availability of digitized corpora of spoken language and inexpensive computing power, the study of language as it is actually used has been receiving more attention. Several of the ideas of usage-based linguists have particular implications for the study of sound systems. These include the notion that experience with categories of sound affects their representation: the more experience the easier the access. Closely related are the ideas that what we know about categorization generally applies to phonological structures (see Rosch 1978, of course). Further, there is no firm separation of language structures and the rules that are applied to them—data structures and algorithms in the language of computer science—as in the formalist tradition (Chomsky and Halle 1968; Pinker 1999), but, rather, linguistic properties emerge from the complex interplay of particular languages and their use, just as do purely biological systems. In fact, in this view, language emerges from repeatedly applying underlying and general cognitive mechanisms (Bybee, 2010). Finally, and more generally, a correct formal characterization of language, individually or collectively, may not be possible and even if it were, the formalism itself does not constitute an explanation of the phenomenon under investigation. Rather, as Bybee and McClelland argue (2005), formalisms describe linguistic regularities that result from the normal process of language use and adaptation.

Hume's Model of Metathesis

Elizabeth Hume's (2004) study of metathesis is an especially nice example of the application of usage-based techniques to a phenomenon that has puzzled linguists for many years. (All examples of metathesis in this paper are taken from Hume). Hume defines metathesis as "the process whereby in certain languages the expected linear ordering of sounds is reversed under certain conditions.

Thus, in a string of sounds where we would expect the ordering to be …xy…, we find instead …yx…" (p. 203). For example, in recent American usage, the word *chipotle,* can frequently be heard as *chipolte*, where /t/ and /l/ are shifted. A very similar kind of metathesis occurs in binyan 5 of perfective verbs in modern Hebrew. When the /-t-/ indicating the binyan 5 morpheme is followed by a stem initial strident (/s/ or /z/, for example), the morpheme and the strident shift expected positions. Thus we have *hitnakem* ("he took revenge") and *hidbalet* ("he became prominent") but, also, *histader* ("he got organized") and hizdaken ("he grew old").

Perhaps the most perplexing element is that a pattern of sounds occurring in one order in language A can occur in the opposite order in language B. Consider examples drawn from Hungarian and Pawnee. In certain Hungarian forms, glottals that precede approximants surface as approximants preceding glottals (/h/ + /r/, in this case, becomes /r/ + /h/). Thus the dative *tehernek* ("load") becomes in the plural *terhek.* In Pawnee, just the opposite occurs. The expected ordering /ti-ir-hissask-kus/ becomes *tihrisasku,* with the glottal appearing before the approximant. According to Hume, this led metathesis to be analyzed as a phenomenon that is irregular, found in child language, the result of performance errors, or simply the result of language change.

In fact, implicit in her discussion, though distinctly underplayed, is that metathesis leads to permanent language change. That is, metathesis is a diachronic phenomenon. This raises metathetic change from a mere curiosity whose regularities can be described to an element of language change. And, as Joan Bybee, a leading figure in the usage-based camp reminds us in her recent book, "nothing in linguistics makes any sense except in light of language change" (Bybee 2010, p. 10). Although the pronunciation of */chipotle/* as */chipolte/,* not simply within a linguistic generation but within a single speaker, can be accounted for by her model, Hume's work becomes really interesting when it tries to account for what was once a puzzling aspect of linguistic change. How, for instance, did the expected /hitsader/ in Modern Hebrew become /histader/? Though diachronic processes are not her primary interest, Hume's account of metathesis can be reframed in evolutionary terms. What any naturally selective process needs is an initial state, an environment that favors certain forms over others, and an output. Hume's work provides all three. The initial state, of course, is "the expected linear ordering of sounds." The output is the reverse ordering. The "certain conditions" correspond to the phonological environment that favors some forms over others.

Hume argues that metathesis requires two conditions:
- An indeterminate speech signal
- An output that conforms to existing patterns in the language.

This is another way of saying that if I don't quite understand what you just said, I'll interpret in light of what I already know. My reinterpretation, of course, will be in the context of what I know best, namely the most frequent sounds in my lexicon. In evolutionary terms, an indeterminate speech signal is one that is not optimally suited to its environment, the "existing patterns of the language." It is important here to clarify a common misconception about natural selection. It does not claim that a given organism is optimized, that it manifests the best possible arrangement of parts. The theory does claim that differential reproduction allows an organism which is better adapted to a specific and limited environment to produce more offspring than one that is not. So, biology is neither random nor goal-directed. Hume makes a similar point about metathesis: "the goal of metathesis is not to improve the overall psychoacoustic (i.e., universal) cues of a sequence, but rather conforming to the patterns of usage of a given language is key" (p. 225). These two ideas, that frequency of use plays a role in language development—see especially, Bybee 2010)—and that metathesis can be reframed as an emergent phenomenon, are the ideas that interest us most and that put Hume's account squarely within the usage-based camp.

Emergentist Models of Language

The view that language is emergent, that it is, in fact, a complex adaptive system, has received attention in recent years. One of the earliest accounts is Lindblom's 1984 attempt to select "with the aid of a *self-organizing model* a 'phonological structure'" [emphasis in the original]. In fact, a snippet from that article, "DERIVE LANGUAGE FROM NONLANGUAGE!," has been used recently used as a summary of the goals of usage-based linguistics (Diessel 2011). More recently, Ke and Holland (2006) note that there are two main approaches to the investigation of language origins. First, there are nativist accounts of language competence and performance that concentrate on cognitive mechanisms and their biological underpinnings. Then there are empirical accounts that concentrate on social structures and patterns of linguistic transmission. In the latter, "language could have evolved from simple communication systems through generations of learning and cultural transmission, without new biological mutations specific to language. While the human species may have evolved to be capable of learning and using language, it is more important to recognize that language itself has evolved to learnable for humans" (Ke and Holland 2006. p. 693).

Andrew Wedel (2005) offers a nice analogy. It seems unreasonable to assert that one's ability to hold a fork is genetically encoded in any precise fashion, despite that fact that humans, as far as is known, are the only species to use them. On the other hand, the manner of fork-holding is culturally transmitted within genetically-encoded parameters, namely four fingers and an opposable thumb. We might even become better fork-holders over time, as our forks evolve to fit our gifts. This notion, that linguistic transmission occurs within species-specific parameters, is captured in the emergentist paradigm. As Ellis put it (cited in Ke and Holland, 2006, p. 694), language acquisition can be explained by "simple learning mechanisms, operating in and across the human systems for perception, motor-action,

and cognition as they are exposed to language data as part of that communicatively-rich human social environment by an organism eager to exploit the functionality of language" (Ellis 1998, p. 657).

Both Holland and Ke (2006) and Holland (2005) situate their work within the tradition of agent-based and complex adaptive systems. Holland—the original developer of the genetic algorithm (Holland 1975)—describes his own efforts to model language acquisition as a complex adaptive system. He uses the phrase "adaptive agent" to describe an individual collection of linguistic rules that communicates with what appears to be a linguistic environment. Some of these agents have a better fit with the environment than others. These survive to evolve still better rules.

Though these accounts are persuasive enough, the real question to be addressed is what one gets after one creates a software model of larger system. O'Reilly and Munakata (2000) make an especially persuasive argument for why one might want to model cognitive processes, the most important piece of which for our own work is that models force investigators to be explicit about their theories. It is one thing to describe a process. It is quite another to describe it with the precision necessary to run it on a computer. Thus Hume draws on Ohala's (1993) observation that certain categories of sound, glottals and liquids for example, (i.e., the closure of the glottis in *bitten* and /r/, see Ladefoged 2006) have "stretched out features" that can bleed over into adjacent sounds causing indeterminancy (Hume, 2004, p. 219). To construct a computer model, we would have to know *how* stretched out. Glottals have cues that are certainly longer than the release bursts of stops (/b/ for example). But *how much* longer? An empirical approach suggests itself immediately: conduct experiments. Another approach, the one implicit in emergentist theory, is to build a model and adjust its parameters until its inputs and outputs conform to the data. In a nutshell, this is what guides our efforts.

The Genetic Algorithm

The Genetic Algorithm (GA) is an optimization method based loosely on the idea of natural selection. Individual members of a species who are better adapted to a given environment reproduce more successfully and so pass their adaptations on to their offspring. Over time, individuals possessing the adaptation form interbreeding populations, that is, a new species. In keeping with the biological metaphor, a candidate solution in a GA is known as a *chromosome*. The chromosome is composed of multiple genes. A collection of chromosomes is a called a *population*. The GA randomly generates an initial population of chromosomes which are then ranked according to a fitness function. One of the truly marvelous things about GA is its wide applicability. We have used it to optimize structural engineering components and are currently applying it to a classic problem in graph theory (Ganzerli, S., De Palma, P. et al., 2003, 2005, 2008). As it happens, both problems are NP-Complete, in effect, computationally intractable (Overbay, S., Ganzerli, S, De

Palma, P., 2006). In practice, of course, this means that those who attempt to solve these problems must be content with good-enough solutions. Though good-enough may not appeal to purists, it is exactly the kind of solution implicit in natural selection: a local adaptation to local constraints, where the structures undergoing change are themselves the product of a recursive sequence of adaptations. This can be expressed quite compactly:

```
GA()
    Initialize(population); //build initial population
    ComputeCost(population); //apply cost function
    Sort(population); //rank population
    while (population has not converged on a good-enough solution)
            Pair(population); //decide which members reproduce
            Mate(population); //exchange characteristics
            Mutate(population); //randomly perturb genes
            Sort(population); //rank population
            TestConvergence(population); //has a new species appeared?
```

The use of the GA to model metathetic change is consistent with Croft's (2000) theory of language change that he calls "utterance selection." In utterance selection, "normal replication is in essence conformity to convention in language use. Altered replication results from the violation of convention in language uses. And selection is essentially the gradual establishment of a convention through language use" (p. 7). In Croft's view, the utterance corresponds to DNA, the replicators to genes, the variants in linguistic structures to alleles. The task in building a model is to find, according to Croft, those mechanisms that cause certain linguistic structures to be favored over others. These are "the causal mechanisms of selection of linguistic structures" (p. 31). Hume's work provides just such a causal mechanism. We show next that this causal mechanism can be modeled with GA.

Metathesis and the GA

Hume describes several kinds of metathesis, all conforming, in one way or another, to her initial claim that metathesis results from indeterminate speech signals processed in terms of frequently occurring sequences of sounds in a given language. The *chipotle/chipolte* example is an instance of this recurring pattern: "a consonant with potentially weak phonetic cues often emerges in a context in which the cues are more robust than they would have been in the expected, yet non-occurring, order" (p. 209). More specifically, stop consonants are easier to perceive in prevocalic position. In fact, over one-third of the metathesis tokens that Hume identifies involve a stop consonant. In the example, [tle] is less favorable in the environment of American English than is [lte]. That is, the stop consonant before the /l/ produces an indeterminate signal for American English speakers, who proceed to shift it to the more frequent pre-vocalic position.

How to represent this process in a GA is the next question. Clearly, we must assign a better fitness, a lower cost, to sequences with pre-vocalic stop consonants than to those with post-vocalic stop consonants. But, somehow,

both signal indeterminacy and token frequency must be made part of this process. Here is our approach:

1. Input an initial population of the base word and the target word. *chipotle* is an example of a base word and *chipolte* is an example of the target word. Our GA works with a total population of 64 words. The relative frequency of the base and target words is a parameter. Thus, we might have one instance of the base and four of the target in the initial population.
2. Generate a random sequence of characters that fill out the population. So, if we seeded the population with one instance of the base and four of the target, our GA would randomly generate fifty-nine character sequences.
3. Assign a fitness value to each of the sequences that comprise the population.
4. Sort, pair, mate, and mutate the population. Sorting is the process of ranking by fitness value. Pairing is the process whereby strings of sounds are collected in two-tuples. As a proof of concept, we adopt a simple approach. The two-lowest cost strings are paired, followed by the next two lowest cost until we have 16 breeding pairs. The remaining 32 strings are discarded to make room for the progeny of our breeding pairs. Mating is the process by which the paired words pass on their genetic composition—their sounds—in the process of generating two new strings of sounds. Mutating is the random shifting of a fixed fraction of the genes in the population. This mimics the action of chemical/biological/radiological mutagens on individuals. For our purposes, it prevents the system from getting stuck in local minima (see Haupt & Haupt 1998).
5. Stop when some predetermined condition is met, else go to step 3.

The cost function in any GA embodies most of the theory being modeled. The other pieces are parameters to the system. The most important of these for us is the relative frequency of the base word and the target, i.e., the initial character sequence and the target of metathetic change respectively. The cost function itself is an attempt to operationalize Hume's model. Except for a few items designed to exclude randomly generated but non-occurring phonetic sequences, it is as follows:

1. A prevocalic stop is more salient than a postvocalic stop. Give a fitness boost to words with prevocalic stops.
2. By observation 1, penalize words with postvocalic stops.
3. Glottals, liquids, glides (/w/, for example) tend to bleed over into adjacent sounds . This is especially true when they follow a stop. Penalize words with glottals, liquids, and glides that follow a stop.
4. A stop followed by a consonant is perceptually weak. Penalize words with stops followed by consonants.
5. A stop followed by a strident is perceptually weak and infrequent. Penalize words with prestrident stops. This rule is what allows our GA to generate the kind of metathetic change found in binyan 5 of perfective verbs in Modern Hebrew (/hitsader/ → /histader/) as well as another instance of English metathesis (/ask/ → /aks/).

Method and Results

Our GA was constructed using Java programming and run under Ubuntu Linux. Its cost function is designed to model, among many other instances, both the *chipotle/chipolte* metathesis as well as binyan 5 of perfective verbs in modern Hebrew, specifically *hitsader/histader*. Every parameter was held constant except the relative frequency of base and target sounds. Since the sounds being modeled occur in the interior of the word in both cases, the strings *potle/polte* and *itsa/ista* functioned as surrogates for the entire words. The population size was set at 64 and the mutation factor set at .5%. For each of 1, 2, and 4 initial *chipotle/hitsader* tokens, the number of *chipotle/histader* tokens began at parity then was doubled three times. So, for instance, if we were working with an initial population of 4 *chipolte* tokens, we would produce results for 4, 8, 16, and 32 *chipolte* tokens. Therefore, there were 12 frequency configurations, four for each set of 1, 2, or 4 *chipotle* tokens. For each of these 12 configurations, we ran the GA 250 times, each run consisting of 250 generations. Along the way, the *chipotle/hitsader* tokens disappeared. The data is summarized in the Tables 1 and 2 below.

Discussion and Future Research

The data illustrates that we were able to design a computational model using the Genetic Algorithm that captures Hume's model of metathetic change. In every one of the 12 frequency configurations, the *chipotle* tokens disappeared from the population within three generations and *hitsader* tokens within two. "Generation," of course is the term used in the GA literature. It is not to be confused with a human generation. Further, within 60 generations, on average, *chipolte* tokens made up an average of 95% of the population. Hebrew metathesis performed even better, with *histader* tokens comprising an average of 97.3% of the population within, on average, 48 generations. At this point, it might be useful to recall Hume's two conditions for metathesis: the speech signal must be indeterminate, and the output must conform to existing patterns in the language. As we indicated with the Hungarian and Pawnee attestations above, metathesis is not just a rule-based phenomenon found in the same form cross-linguistically. Rather, it is intimately tied to existing sound patterns within a language. Said another way, metathesis is a usage-based phenomenon. Our model demonstrates this in terms of a very solid frequency effect. The maximum number of target tokens tends to stabilize more quickly and at a higher percent of the total population as the number of target tokens in the initial population increases. Further, the larger the set, where a set

is defined as the number of base tokens in the initial population, the better the performance. This is illustrated most strongly when we look at data from the first and last element of each configuration; that is when we compare 1:1, 2:2, and 4:4 with 1:8, 2:16, and 4:32. The more frequent the target within the initial population, the more quickly the population stabilizes on the target and at a higher percent of the total population.

Nevertheless, Hume's model is underspecified from an algorithmic/computational standpoint. Though it specifies very clearly what kinds of sounds are potentially vulnerable to metathetic change and in what context, the computational modeler must guess how to weight the various phonetic factors involved and, in particular, to guess at frequency thresholds. We regard our study as a proof of concept. In future work we will build our frequency hypotheses into the rules themselves. For example, instead of simply rewarding strings with a prevocalic stop and penalizing those with a postvocalic stop, we will use transcribed corpora to estimate the frequency of both vulnerable cues and the targets of metathetic change. These frequencies will be used to weight the penalties and rewards, thus making as precise as possible important observations like, "Indeterminancy sets the stage for metathesis, and the knowledge of the sound patterns of one's language influences how the signal is processed and, thus, the order in which the sounds are parsed" (Hume, 2004, pp. 209- 210). Our goal is that by gathering data on vulnerable sounds in corpora of actual speech, we will be able to generate all of the instances of metathesis within a language. This will add weight to Hume's observations and perhaps be useful in accounting for and predicting other types of language change.

Table 1: Chipotle, 250 Runs, 250 Generations Each

Ratio of Base to Target	Generation *Chipotle* Disappeared	Generation *Chipolte* Stabilized	Percent of *Chipolte* Tokens at Stabilization
1:1	3	119	73.4
1:2	3	68	93.7
1:4	3	60	96.8
1:8	2	44	98.4
2:2	3	90	92.1
2:4	3	72	96.8
2:8	2	50	98.4
2:16	2	31	98.4
4:4	3	58	96.8
4:8	2	44	98.4
4:16	2	33	98.4
4:32	1	26	98.4

Table 2: Hitsader, 250 Runs, 250 Generations Each

Ratio of Base to Target	Generation *Hitsader* Disappeared	Generation *Histader* Stabilized	Percent of *Histader* Tokens at Stabilization
1:1	2	79	84.3
1:2	2	65	98.4
1:4	2	55	98.4
1:8	2	39	98.4
2:2	2	59	98.4
2:4	2	47	98.4
2:8	2	43	98.4
2:16	1	33	98.4
4:4	2	49	98.4
4:8	1	43	98.4
4:16	1	32	98.4
4:32	1	23	100

References

Bybee, J. 2001. Phonlogy and Language Use. Cambridge: Cambridge University Press.

Bybee, J. 2010. Langauge, Usage and Cognition. Cambridge: Cambridge University Press.

Bybee, J., and McClelland, J. 2005. Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. The Linguistic Review 22,381-410.

Chomsky, N., and Halle, M. 1968. The Sound Pattern of English. NY: Harper and Row.

Croft, W. 2000. Explaining Language Change: An Evolutionary Approach. Harlow, England: Pearson.

Ellis, N. 1998. Emergentism, Connectionism, and Language Learning. Language Learning 48, 631-634.

Ganzerli, S., & De Palma, Paul. (2008). Genetic Algorithms and Structural Design Using Convex Models of Uncertainty. In Y. Tsompanakis, N. Lagaros, M. Papadrakakis, eds. Structural Design Optimization Considering Uncertainties. London: A.A. Balkema Publishers, A Member of the Taylor and Francis Group.

Ganzerli, S., De Palma, P., Stackle, P., Brown, A. 2005. Info-Gap uncertainty on structural optimization via genetic algorithms. Proceedings of the Ninth International Conference on Structural Safety and Reliability, *Rome.*

Ganzerli, S., De Palma, P., Smith, J., & Burkhart, M. 2003. Efficiency of genetic algorithms for optimal structural design considering convex modes of uncertainty. Proceedings of The Ninth International Conference on Applications of Statistics and Probability in Civil Engineering**,** *San Francisco.*

Haupt, L. & Haupt, S. 1998. Practical Genetic Algorithms. New York: John Wiley and Sons.

Diessel, H. 2011. Review Articles: Language, usage, and cognition. Language 87(4):830-844.

Holland, J. 1975. Adaptation in Natural and Artificial Systems. Ann Arbor: The University of Michigan Press.

Holland, J. 2005. Language acquisition as a complex adaptive system. In Minett, J., Wang, W. eds. Language Acquisition, Change and Emergence: Essays in Evolutionary Linguistics. Hong Kong: City University of Hong Kong Press.

Hume, E. (2004). The Indeterminancy/Attestation Model of Metathesis. Language 80(2): 203-237.

Ke, J., & Holland, J. 2006. Language Origin from an Emergentist Perspective. Applied Linguistics 27(4): 691-716).

Ladefoged, P. 2006. A Course in Phonetics. Boston: Thomson/Wadsworth.

Lindblom, B., MacNeilage, P., & Studdert-Kennedy, M. 1984. Self-organizing processes and the explanation of phonological universals. In Butterworth, B., Comrie, B., Dahl, O. eds. Explanations for Language Universals. New York: Mouton.

O'Reilly, R., & Munakata, Y. 2000. Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain. Cambridge: MIT Press.

Ohala, J. 1993. Sound change as nature's speech perception experiment. Speech and Communication 13.155-161.

Overbay, S., Ganzerli, S., De Palma, P, Brown, A., Stackle, P. 2006. Trusses, NP-Completeness, and Genetic Algorithms. Proceedings of the 17th Analysis and Computation Specialty Conference. St. Louis.

Pinker, S. 1999. Words and Rules. NY: HarperCollins.

Rosch, E. 1978. Principles of Categorization. In E. Rosch and B. Loyd (eds.), Cognition and Categorization, 27-48. Hillsdale, NJ: Lawrence Erlbaum Associates.

Wedel, A. Contrast Maintenance in Language and the Innateness Debate. Retrieved 2/18/2013 from http://dingo.sbs.arizona.edu/~wedel/research/PDF/wedelcontrastsummary.pdf