Probability in Phonological Generalizations: Modeling French Optional
Final Consonants
Author(s): Benjamin K. Bergen
*Proceedings of the Twenty-Sixth Annual Meeting of the Berkeley
Linguistics Society: General Session and Parasession on Aspect* (2000),
pp. 27-39

**Probability in phonological generalizations:
modeling French optional final consonants**

BENJAMIN K. BERGEN
U.C. Berkeley and *International Computer Science Institute*

0.   The problem of conflicting phonological generalizations[*]

The place for statistics is outside, not inside, grammar. (Smith 1997:298)

it is important ... not to mistake complexity for irrelevancy. (Abney 1997:13)

The starting point for this paper is the simple and unoriginal observation that many phonological generalizations are *variable*. By *variable*, I mean that their application is not predictable from phonological properties of the string, but rather depends probabilistically on other factors. In speech perception, phonological generalizations are particularly variable because of dialectal and inter-speaker differences; perception is the perspective taken in this paper.

Four types of variable phonological generalizations are identified in section 1. I outline several existing partial solutions to these types in section 2, from the generative phonology and variable rules traditions. These solutions cannot account for all types of variation, though, and I argue in section 3 that one type they have particular problems with, *interacting extraphonological variation*, is qualitatively indistinguishable from the others.

Section 4 proposes a solution which takes advantage of a knowledge representation model recently developed in Artificial Intelligence known as Belief Networks (BNs–Jensen 1996). Section 5 demonstrates that BNs not only meet descriptive adequacy for interacting extraphonological variation but also that they surpass variable and generative models in learnability and neural plausibility.

1.   The variable generalization problem
1.1   Types of variability
Despite what one might gather from a survey of introductory phonology textbooks, phonological generalizations are not all simple and invariant. And yet it is possible to detect traces of variation in the generative phonological literature, where four types of variation phenomena can be distinguished.

Type 1. **Free** (*Type 1*) variation is occasionally mentioned, but infrequently accorded any import. At least, this was the case before Optimality Theory

(OT–Prince & Smolensky 1993) took the phonological stage. To the extent that any variation is truly free of social correles, it can be dealt with much more parsimoniously in OT than in previous generative models, and very interesting quantitative work has been done in this direction (e.g. Antilla 1997).

Type 2. **Predictable** phonological (*Type 2*) variation is much more commonly discussed in the generative literature because it is thought to be rule-governed. *Type 2* variation can be defined as the variable application of a phonological generalization when certain other phonological generalizations are present. A central example of this type of variation is phonological *opacity* (Kiparsky 1973).

Type 3. Not all non-phonological variation (perhaps none of it) is truly free. Rather, certain variable phonological generalizations, like the deletion of a word-final *t* or *d* in English, correlate with social factors. Research on Variable Rules (VRs), which started in the late 1960s (Labov 1966) and has continued into the present (c.f. Fasold 1996 for an overview) stems from this very observation.

Sudents of sociolinguistic variability have closely inspected **independent** (*Type 3*) variable generalizations. An example of this class is the case of French optional final consonants (Verluyten & Hendrickx 1987, Stammerjohann 1976). These are a class of about 50 words whose final consonants can be pronounced or elided, depending on social and phonological environment; they include *août* 'August', *cinq* 'five', *fait* 'fact', *but* 'goal', and *ananas* 'pineapple'. The most significant conditioning factors for the rule are properties of the following segment and the speaker's age, gender, nationality and social class (1).

(1) Sociolinguistic variables & French optional Cs (Verluyten & Hendrickx 1987)

	Age		Nationality		Gender		Social Class	
	Young	*Old*	*French*	*Belg.*	*F*	*M*	*Low*	*High*
% produced	65	52	72	47	61	57	65	52

None of these conditioning factors *determines* whether the final consonants will appear in any strong sense. Instead, particular values for each variable increase the statistical likelihood that the consonant will occur. For example, the categorization of the segment following the optional final consonant can affect these values: optional final consonants are missing before consonants with a 64.8% probability, and before vowels at 48.8%. Moreover, these contraints do not interact; the contribution of each can be assessed independent of the values of any of the others. This distinguishes *Type 3* (independent) variability from *Type 4* (interacting) variability.

Type 4. **Interacting** (*Type 4*) variation is nicely exemplified by the problem of word length and grammatical class in French liaison (Tranel 1981). French liaison consonants are a class of word-final consonants that can be pronounced or not, depending again on a set of factors. These factors belong to various domains: surface phonological (e.g. properties of other realized segments), syntactic (e.g. constituency and grammatical class), morphological (e.g. expressive/semantic status), social and sociolinguistic (e.g. register, style, speech rate, and dialect), and others. Again, they contribute non-deterministically to the string's realization.

Crucially, two of these factors interact: word length and grammatical class. In

other words, shorter liaison words make the expression of a liaison consonant more likely, but only if those words are prepositions or modifying adjectives, and not nouns or non-modifying adjectives (de Jong 1989). Thus we say that these two effects are not *independent*, but rather *interacting*. The theoretical ramifications are addressed in section 3, but in sum, existing generative models and variable rule analysis cannot deal with generalizations of this type.

1.2. Evidence for variability

One explanation for the poverty of discussion of variation in the mainstream phonological literature is that it calls for a particular methodology that is not the norm. Variation benefits from corpus studies in a way that invariance does not for three main reasons. First, since correlates of a variable phonological generalization can be of various types, the set of possible contributing factors can become too large to be elicited or introspected in a reasonable time. Second, the very fact that the generalizations are not deterministic means that a large number of tokens need to be studied for results to be statistically significant. Finally, some variability is closely tied to extralinguistic attributes of the speaker, in which case a single speaker is not a representative source of evidence.

A second explanation is the belief in a fundamental schism among the four types of variability discussed above. Where exactly to draw that line is unclear, but opacity (*Type 2*) has been widely addressed by mainstream phonological models, while interacting extraphonological variation (*Type 4*) has not. In contrast, two arguments can be made for treating all four variation types uniformly.

First, the kind of evidence that a language-perceiver or analyst can collect for all types is qualitatively similar. All cases are characterized by measurably different phonological forms with identical or related denotations, where phonological generalizations about the distribution of those forms cannot be made on the basis of their surface co-occurrents. The types differ in whether the probabilities of the generalizations correlate with factors outside the string, and whether contributing factors interact. Opacity, e.g., can be seen as interaction between morphological and phonological generalizations.

The other argument for treating variability in a unified manner is that no clean line can be drawn between a purely phonological or purely grammatical variability on the one hand and extraphonological or extragrammatical variability on the other. Rather, different contributing factors from different modes interact. For example, properties of the following segment interact with lexically-specified final-consonant probabilities in French liaison (Bergen ms).

A final explanation for the absence of invariability in phonological studies is the *a priori* belief that phonological systems are inherently deterministic. The next section explores the ramifications and limitations of this notion, as well as the Variable Rule model, which is a step in the direction of representing uncertainty.

2. Dealing with variable generalization

Most phonological models are characterized by an underlying assumption, the *invariance hypothesis*, which holds that the elements of phonological generalizations are invariant. In particular, invariance is presupposed for the

object of a generalization, its environment, and its content. As Labov (1997) notes, the principle aim of linguists seems to be the reconstruction of theoretical invariance on the basis of variable data. Historically and methodologically, it is precisely the failure of this procedure that leads to the study of variability. But how is an invariant model to deal with statistically-variable data?

The study of *Type 1* (free) variable generalizations is still in its infancy in invariant generative phonology. While rule-ordering approaches have essentially nothing to say about free variation, several different means have recently been proposed for making precise distribution predictions using OT. These include the invention of probabilistic ranking (Boersma & Hayes ms), relative unranking (Anttila 1997), and floating of constraints (Nagy & Reynolds 1997).

Type 2 (predictable) phonological generalizations do not hold in a surface-true manner, but due to regular historical developments—and sometimes also to synchronic factors—they do display certain regularities. These regularities are integrated into invariant models where rules can be intrisically ordered (Kiparsky 1973), where constraints "gang up" (Kirchner 1996) or where forms can be selected by the analyst to win on grounds of "sympathy" (McCarthy 1997).

However, *Type 3* (extraphonological independent) variability seems to be best addressed by Variable Rules (VRs). VR analysis (Labov 1966) treats socially-correlated variation with the very simple but theoretically significant addition of quantitative weightings to SPE-style generative rules. The most discussed example of this type of analysis treats the problem of English word-final *t/d* deletion (e.g. Guy 1991). The application of *t/d* deletion, although apparently active to some extent in all dialects of English, has been found to correlate with certain social attributes of the speaker, such as age, social class, and gender. Linguistic environments favoring the application of this rule are two preceding consonants, an unstressed final syllable, and a following sonorous segment, i.a.

VRs take the following form, where angled brackets designate variable contexts, those that affect the application of the rule probabilistically given a socially-determined input probability, and square brackets are invariant (2). VRs like this one are derived from the statistical evaluation of the probabilistic contribution of a set of independent factors to the application of the rule.

(2) t, d → <Ø> / <-stress> <+cons> [+cons] _ <+son>

Neither generative nor variable models is able to address *Type 4* (interacting) variability. The next section describes failings of each model in accounting for variability.

3.   Problems with existing models

The best-known critiques of derivational models are psychological and neural plausibility arguments (c.f. the papers in Goldsmith 1993 or Lima et al 1994). One of the most convincing is a processing limit argument. Adapting Price's (1996) demonstration, let us imagine a truly serial perceptual (or production) system, in which each of six modules makes its calculations, then passes its results to the subsequent module. (For argument's sake, these might be phonetic, phonological, morphological, syntactic, semantic, and pragmatic levels.) Given that no module

can ever be 100% certain of its assessments (due to variability ranging from the phonetic to the semantic realms), an ideal system might display 95% accuracy per module. But giving these nearly perfect calculations, we could only be confident about the product of our elaborate derivation to an accuracy of only 73%. In vivo production and recognition systems work worlds better than this.

Strong seriality therefore calls into question the plausibility of VRs and certain generative models, independent of problems they display with variability.

3.1. Invariant models

The invariance itself of invariant phonological models poses problems when it comes to learning. How can an invariant representation be constructed on the basis of variable input? The simplest, and perhaps the most theoretically-charged, answer is to posit innate and/or universal pre-phonological generalizations (read: *parameters* or *constraints*). Hoping not to get sidetracked by the innateness question, I would simply like to point out that "learning systems" based on innateness are not accounts of the <u>learning</u> of invariant structures from variable input, but rather accounts of the <u>reorganization</u> of invariant generalizations on the basis of variable input (e.g. Tesar & Smolensky 1996, Boersma & Hayes ms). Moreover, innateness is not the only solution; there exist psychologically-plausible accounts of concept formation that depend on exclusively variable inputs, such as prototype (Rosch 1978) and exemplar models (Johnson 1997). To the extent that learning of variable concepts is possible, arguments for innateness on the basis of the poverty of the stimulus are unjustified, as thus so are models based upon the reorganization of invariant, innate constraints.

Moreover, the extension of invariant models to account for socially-correlated variation (*Types 3* and *4*) may violate their basic tenets. The circular reasoning might go something like this: Grammar is different from usage-based linguistic knowledge in part in that it is invariant and algebraic; it is a separate module that does not interact with extralinguistic knowledge; social factors cannot enter into a model of grammar because they introduce uncertainty and language-external constraints ("There is no motivation for tying variation to rules of grammar" (Fasold 1996:91)).

3.2. Variable Rules (VRs)

Certain long-standing critiques of VRs have enjoyed wide success, and VRs have become somewhat hard to come by, as summed up nicely by the title of Fasold's (1996) paper, "The quiet demise of variable rules". Attacks have focused on both empirical and theoretical commitments. First, I cite some empirical claims about language for which VRs have been criticized :

- **They are unable to account for the statistically interacting properties of social contexts and deny this property for linguistic context.** This aspect is problematic since social and linguistic factors display interacting properties (de Jong 1989).
- **They claim different languages or varieties stand on a continuum.** On the contrary, language users often recognize clear-cut boundaries between speech communities.

- **They presuppose variables in VRs to have no effect on meaning.** The semantics of the utterance must not be affected by a choice of alternate.
- **They disallow conscious manipulation of linguistic variable, as with speech style** (Dittmar 1996). Social meaning is restricted to phonological and morphological variation. This entails that speakers cannot make linguistic choices to convey social meaning.

Variable rules have also been attacked for theoretical properties:

- **Their phonological representations inherit SPE's flat structure.** Hierarchical organization of phonological structures was one of the most significant advances of recent phonological theory (Goldsmith 1990). VRs have not been extended to hierarchical phonology, probably because of the enormous structural and computational problems of such a move.
- **Their quantitative values are essentially arbitrary, i.e. non-explanatory**, except that they do predict relative ranking of environments (but see Guy's (1991) claim to contrary, which appears to suggest ratios on the basis of a lexical phonology approach).[1]
- **They don't actually integrate probabilities with a grammar.** VRs provide interface points, but the probabilistic mechanisms that would have to be posited to deal with variation are absent.
- **They can cover up categorical behavior by individuals or subgroups.** Bickerton (1971) observed that statistical generalization can lead to the belief that individual behavior is identical to group behavior.

Both VR and invariant models are problematic in general, and are unable deal with uncertainty of various types, in particular, *Type 4* variability.

4.   Fighting uncertainty with uncertainty

I propose an alternative solution to the variable surface generalization problem which does not make use of problematic ranking or ordering and additionally allows the complex combination of multiple modes of probabilistically interacting information. The proposed solution posits that phonological knowledge itself is not invariant but rather probabilistic. *Type 1* (free) variation is dealt with by assigning a prior probability to the variable phonological generalization. *Type 2* (predictable) variation involves the near-categorical, but nonetheless probabilistic, contribution of phonological and morphological factors. *Type 3* (independent) variation involves the probabilistic assessment of external (and phonological) factors. *Type 4* (interacting) variation is treated with interacting probabilistic external and phonological factors. A single, probabilistic mechanism can serve to unify these types of knowledge.

---

[1] The same critique could be leveled at generative grammar in general. Differences in language-specific constraint rankings are entirely non-explanatory, even if they can be harnessed to give specific quantitative predictions about realization (as in Anttila 1997).

4.1.   Support for probability in cognition

The proposal that phonological knowledge is fundamentally probabilistic finds independent support in both general cognitive processing properties and specific aspects of "creative" language use.

Schematic conceptual knowledge (that is, abstraction over specific perceptual instances) is probabilistic rather than invariant. Several lines of fruitful research dedicated to understanding conceptual representations have formulated this observation in differing terms: fuzziness, prototype effects, and gradedness in categories (Lakoff 1987), but the substance of the models is similar. We can thus expect phonological abstractions to be probabilistic as well.

Second, "creative" uses of language (the production or recognition of novel forms) and historical developments (Bybee and Slobin 1982) display probabilities in the same way as "uncreative" language does. By the same token, linguistic judgements (a sort of "creative" linguistic endeavor) are similarly subject to probabilistic assessments (Hayes ms, Bender 2000). Along with the probabilistic nature of corpus data, this suggests that empirical evidence for creative language use is in essence probabilistic.

A model which aims to extract multi-leveled probabilistic information from the phonological signal, and also to adapt its production according to multi-modal factors impacting phonology, must encode this knowledge. Until recently, however, the computational demands of this problem were too great. Belief Networks are shown in the next section to be appropriate for modeling phonological generalizations, specifically for the set of interactions responsible for the optional final consonant phenomena described in section 1.1 above.

4.2.   Belief Networks (BNs)

Belief Networks (BNs) are a concise and powerful computational representation of uncertain propositional knowledge (Jensen 1996). Specifically, BNs consist of (1) a set of *nodes* representing propositions or variables, each of which has (2) a set of possible *values*, (3) *links* between causally-related propositions (where causation can be interpreted either ontologically or epistemically), and (4) *conditional distributions*, specifying the probability of each value of every node given a value assignment to its parents. Using a probability theory, inferences can be made about the probability of the value of any node given any (set of) observed values for any other nodes. Another appealing property of BNs for many large-scale problems like the present one is that given certain independence assumptions, the set of conditional distributions is much more succinct than a complete joint distribution for all the variables of the system.
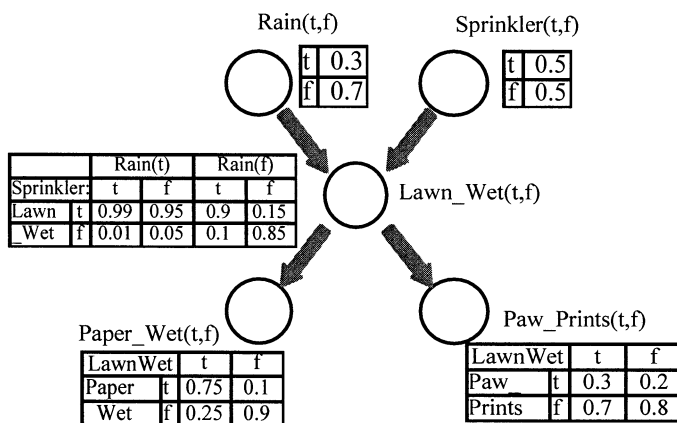
In a simple example, five propositions, each with multiple possible values, are represented by nodes (circles) in (3). The node representing the proposition Rain(t,f) stands in a causal relation to Lawn_Wet(t,f), as indicated by the link connecting the two. Causality is indicated by the unidirectionality of the link; Rain(t) causes Wet_Lawn(t), and not the reverse. Each node is associated with a *conditional distribution table*. Orphan nodes, those with no parents like Rain(t,f) and Sprinkler(t,f), have simple prior distributions that express solely the probability of each of their values. In the example in (3), there is a 0.3 likelihood

that Rain will take the value *true*, and a 0.5 chance that Sprinkler will be *true*. The sum probability for all the values of a proposition is always 1.

The relationship between two causally-linked propositions is encoded in the conditional distribution of the downstream node. Lawn_Wet, for example, has two parents, and since each of them has two possible values, the probability of each of its two values is specified for the four (2^2) possible causal states, thus giving eight (2^3) possible configurations. If we know Rain to be *true*, and Sprinkler to be *false*, then the probability of Lawn_Wet(t) in this example is 0.95, while if Rain is *false* and Sprinkler is *true*, then Lawn_Wet(f) has a 0.1 probability.

(3) Simple Belief Network

Rain(t,f)

t	0.3
f	0.7

Sprinkler(t,f)

t	0.5
f	0.5

Lawn_Wet(t,f)

Sprinkler:	Rain(t)		Rain(f)	
	t	f	t	f
Lawn	t	0.99	0.95	0.9	0.15
Wet	f	0.01	0.05	0.1	0.85

Paper_Wet(t,f)

LawnWet		t	f
Paper	t	0.75	0.1
Wet	f	0.25	0.9

Paw_Prints(t,f)

| LawnWet		t	f
Paw	t	0.3	0.2
Prints	f	0.7	0.8

But the real interest of BNs lies not just in their representational power but more importantly in their inferential power. First, given a network and full set of conditional distribution tables as in (3), beliefs about the various propositions are propagated to produce the *unconditional probability* of each proposition, the prior probabilities of its values given the network. Alternatively, inference can be made given observations about the values of propositions. The second sort of inference, *diagnostic inference*, involves the propagation of evidence from an observed effect (a child) to an unobserved cause (a parent). For example, given that we observe Paw_Prints to be *true*, we might ask what the probability is of Rain or Lawn_Wet being *true*. Third, *causal inference* involves the prediction of effects given that values for causes are observed. In the case of (3), we might observe Sprinkler to be *false* and then let the inference algorithm determine the probability of Paper_Wet also being *false*. Finally, hybrid types of inference are possible: for example, what is the probability of Lawn_Wet being *true* if Rain and Paw_Prints are both *true*?

4.3. A Belief Net model for optional consonants
Assuming for the purpose of exposition that the only factors contributing to the

occurrence of a final consonant in French are the ones enumerated in (1) plus the nature of the first segment of the following word, the BN in (4) can be constructed. Each node represents a variable, and Gender, Class, Age, Nationality, and Following_V are assumed to have only two possible values each to keep the network simple, although continuous values are also allowed in BNs. By the same token, the lexicon is compressed into only three words, each of which represents a word class; *ananas* 'pineapple' has a frequently omitted final consonant, the final consonant of *cinq* 'five' is rarely deleted, and the final consonant of *coq* 'rooster' is always pronounced. Since there are no interactions between individual social factors and any other factors, their influence is summarized under Soc_Factors, which is what is known as a *hidden node*. That is, evidence for the values of this node cannot be directly extracted from the environment itself, but rather must be inferred.

(4)   Belief Network for optional consonants
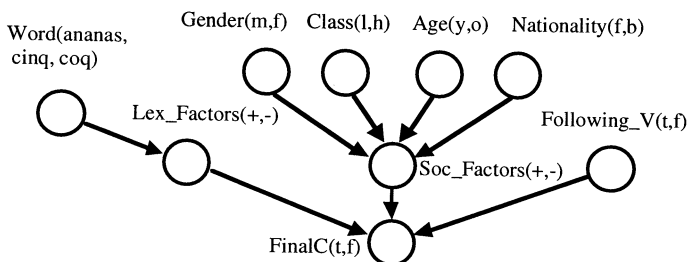
In the network in 4, given the probabilities of co-occurrence of variables taken from Verluyten and Hendrickx (1987), we can perform inference tests. First, let us consider the unconditional (prior) probabilities for each of the nodes, in (5a). Word(coq) is significantly more likely than its lexical competitors, since the non-optional class it stands for is the most common in the language. Final_C(t) is significantly positive as a consequence of this; all other nodes are at chance. Next, we can assess probabilities given observed values for a subset of the nodes. Given, for example, Word(ananas), that is, given that the word considered is *ananas*, the probabilities of all visible nodes are 0.5 for each state, except for FinalC, whose probability of *true* is 0.67 (5b – note that observed values are designated by ***bold italic***). For Word(ananas) and Following_V(f), the only difference is that, as expected, the probability of FinalC(t) rises to 0.73 (5c). To test the significance of FollV, (5d) shows the results of Word(ananas) and FollV(t). The result of setting all the social factors plus Following_V to the values most conducive to suppressing a final consonant is shown in (5e), while the reverse is shown in (5f).

35

(5) Probabilities of nodes in (4) for Word(ananas) and Following V(t)

	Word			FinalC		FollV		Age		Nation		Gender		S.Class	
	Ananas	*Cinq*	*Coq*	*T*	*F*	*T*	*F*	*Y*	*O*	*F*	*B*	*F*	*M*	*L*	*H*
a	.05	.05	.9	.96	.04	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5
b	*1*	*0*	*0*	.67	.23	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5
c	*1*	*0*	*0*	.73	.27	*0*	*1*	.5	.5	.5	.5	.5	.5	.5	.5
d	*1*	*0*	*0*	.54	.46	*1*	*0*	.5	.5	.5	.5	.5	.5	.5	.5
e	*1*	*0*	*0*	.41	.59	*0*	*1*	*0*	*1*	*0*	*1*	*0*	*1*	*0*	*1*
f	*1*	*0*	*1*	.8	.2	*1*	*0*	*1*	*0*	*1*	*0*	*1*	*0*	*1*	*0*
g	*1*	*0*	*0*	*1*	*0*	.58	.42	.52	.48	.53	.47	.52	.48	.52	.48
h	*1*	*0*	*0*	*0*	*1*	.37	.63	.46	.54	.46	.54	.47	.53	.46	.54

Perceptual conclusions can be drawn through inference from an observed input. For example, given that the word *ananas* was identified and produced with or without a final -*s*, the values for all other variables are assessed to be as shown in (5g) and (5h), respectively.

The use of a model like the one described above for speech recognition should be clear; if values for these propositions can be extracted from speech input, they can help to predict future linguistic behavior for disambiguation. BN models constitute dynamic models in this respect. Additionally, in this model, interacting constraints (*Type 4)* are encoded just like independent constraints: the conditional distribution tables identify the contribution of each parent given each value of every other parent. Other properties of the model are discussed below.

5. Properties of this model

The probabilities of the interactions between generalizations can be easily learned in a Bayesian model from surface-true generalizations. Moreover, this computational mechanism displays neural properties, making it cleanly groundable in a biologically plausible model.

Learning the conditional probabilities of a BN involves the relatively simple statistical extraction of distributions of co-occurrent proposition states from evidence. However, structure induction, the construction of a network (and its architecture) from data, is harder. Several methods exist for coercing the right kinds of structure to emerge, and I will only name some here: entropy methods, score metrics, simulated annealing, and genetic algorithms (Jordan 1998).

Not only are BNs learnable from variable data, but they display properties that make them particularly neurally plausible. First, all of a node's information is stored locally: everything a node needs to compute the effects of events elsewhere in the network is available in the node representation itself. Neurons behave in exactly this way. Inference in BNs is performed through the propagation of beliefs from one node to another, in way similar to the propagation of activation in neural systems. Finally, the result of inference in a BN is a probabilistic result, which is analogous to the graded output of a neuron or batch of neurons responding to an input over time.

In recent work, Wendelken and Shastri (ms) have made great strides towards a theory of BN inference at the structured connectionist level (Feldman 1988). They

demonstrate that a class of BNs to which the ones presented above belong can be reduced to structured connectionist systems with very interesting properties: conditional probabilities can be represented as link weights, node states as activation states, and inference propagation as spreading activation. Moreover, they have found that learning can be done in a Hebbian manner in such systems. Hebbian learning is responsible for associative learning – it is the strengthening of connections that fire together. In other words, BNs can be learned at the neural level by the simplest (and oldest) known form of associative neural reorganization.

6.   Conclusion

The argument presented above would benefit from advances on two fronts. First, the demonstration that phonology-internal and phonology-external constraints interact probabilistically is presented elsewhere (Bergen ms). Second, the model must be extended to deal with *Type 2* (predictable) variation. Such a model, following along the same lines as the work presented here, would involve knowledge of variation in forms that are morphologically-related to a given variable form helping a language user predict properties. In other words, knowing that a form has a certain disposition relative to a variable generalization can allow a language learner (and for the same reason a language user) to predict behavior of morphologically-related forms relative to other variable generalizations.

I have tried to argue the position that variation of linguistic generalizations compels us to integrate probability into phonological models, a position in significant accord with usage-based models of phonology, as proposed by Bybee (1999) and Kemmer and Israel (1994). The twentieth century was marked by the introduction of probabilistic notions of causation into physics, allowing some of the greatest technological advances of our time. There is no reason to think that the twenty-first century will not see the same intellectual explosion in language science, as long as we allow ourselves the privilege of thinking probabilistically.

References

Abney, Steven. 1996. Statistical methods and linguistics. In Judith Klavans and Philip Resnik (eds.) *The balancing act: combining symbolic and statistical approaches to language*. Cambridge: MIT Press: 1-26.

Anttila, Arto. 1997. Deriving Variation from Grammar. In Frans Hinskens, Roeland van Hout, and W. Leo Wetzels (eds.) *Variation, Change, and Phonological Theory*. Amsterdam: John Benjamins: 35-68.

Bender, Emily. 2000. Non-categorical constraints in perception. Paper presented at the 24th Annual Penn Linguistics Colloquium, Philadelphia.

Bergen, Benjamin K. ms. The probabilistic combination of phonological and extra-phonological constraints: dimensions of French liaison.

Bickerton, Derek. 1971. Inherent variability and variable rules. *Foundations of Language* 7: 457-92.

Boersma, Paul and Bruce Hayes. ms. Empirical tests of the gradual learning algorithm.

Bybee, Joan. 1999. Usage-Based Phonology. In Darnell et al (eds.) Functionalism & Formalism in Linguistics. Philadelphia:John Benjamins:211-242.

Bybee, Joan L. and Dan I. Slobin. 1982. Rules and Schemas in the development and use of the English past tense. *Language* **58**: 265-289.

Dittmar, Norbert. 1996. Descriptive and explanatory power of rules in sociolinguistics. In Rajendra Singh (ed.) *Towards a critical sociolinguistics*. Amsterdam: John Benjamins: 115-150.

Feldman, Jerome. 1988. Connectionist Representation of Concepts. In D. Waltz and J. Feldman (eds.) *Connectionist Models and Their Applications*. Ablex.

Fasold, Ralph. 1996. The quiet demise of variable rules. In Rajendra Singh (ed.) *Towards a critical sociolinguistics*. Amsterdam: John Benjamins: 79-98.

Goldsmith, John. 1990. *Autosegmental and Metrical Phonology*. Oxford: Basil Blackwell.

Goldsmith, John. 1993. Harmonic Phonology. In J. Goldsmith (ed.) *The Last Phonological Rule: Reflections on Constraints and Derivations*. Chicago: University of Chicago Press.

Guy, Gregory. 1991. Explanation in variable phonology: An exponetial model of morphological constraints. *Language Variation and Change* **3**: 1-22.

Hayes, Bruce. ms. Gradient Well-Formedness in Optimality Theory.

Jensen, Finn V. 1996. *An introduction to Bayesian networks.* New York: Springer.

Johnson, Keith. 1997. Speech recognition without speaker normalization: an exemplar model. In Keith Johnson and John Mullennix (eds.) *Talker Variability in Speech Processing*. San Diego: Academic Press: 145-165.

de Jong, Daan. 1989. A multivariate analysis of French Liaison. In M. Schouten and van Reenen (eds.) *New Methods in Dialectology*. Dordrecht:Foris:19-34.

Jordan, Michael I. (ed.). 1998. *Learning in Graphical Models*. Kluwer Academic.

Kemmer, Suzanne and Michael Israel. 1994. Variation and the Usage-Based Model. *CLS 30*.

Kiparsky, Paul. 1973. Phonological representations. In Osamu Fujimura (ed.) *Three Dimensions of Linguistic Theory.* Tokyo: Taikusha:1-136.

Kirchner, Robert. 1996. Synchronic chain shifts in Optimality Theory. *Linguistic Inquiry* **27**: 341-350.

Labov, William. 1966. *The social stratification of English in New York City.* Washington D.C.: Center for applied Linguistics.

Labov, William. 1997. Resyllabification. In Frans Hinskens, Roeland van Hout, and W. Leo Wetzels (eds.) *Variation, Change, and Phonological Theory.* Amsterdams: John Benjamins: 145-180.

Lakoff, George. 1987. *Women, Fire, and Dangerous Things*. Chicago: University of Chicago Press.

Lima, Susan et al (eds.). 1994. *The Reality of Linguistic Rules*. Philadelphia: John Benjamins.

McCarthy, John. 1997. Sympathy and Phonological Opacity. Rutgers Optimality Archive 252-03981.

Prince, Alan and Paul Smolensky. 1993. Optimality Theory: Constraint Interaction in Generative Grammar. RuCCS Technical Report #2. Piscateway: Rutgers University Center for Cognitive Science.

Nagy, Naomi and Bill Reynolds. 1997. Optimality Theory and word-final deletion in Faetar. *Language Variation and Change* **9**: 37-55.

Price, Patti. 1996. Combining linguistic with statistical methods in automatic speech understanding. In Judith Klavans and Philip Resnik (eds.) *The balancing act: combining symbolic and statistical approaches to language.* Cambridge: MIT Press: 119-134.

Rosch, Eleanor. 1978. Principles of categorization. In Rosch and Lloyd (eds.) *Cognition and Categorization.* Hillsdale, NJ:Lawrence Erlbaum.

Smith, Norval. 1997. Shrinking and hopping vowels in Northern Cape York. In Frans Hinskens, Roeland van Hout, and W. Leo Wetzels (eds.) *Variation, Change, and Phonological Theory.* Amsterdam: John Benjamins: 267-300.

Stammerjohann, Harro. 1976. Zur Aussprache französischer Endkonsonanten: août, but, cinq, usw. *Die Neueren Sprachen* **75**(5): 489-502.

Tesar, Bruce and Paul Smolensky. 1996. Learnability in Optimality Theory. Technical Report JHU-CogSci-96-3. http://www.cogsci.jhu.edu/TechReports/

Tranel, Bernard. 1981. *Concreteness in Generative Phonology: Evidence from French.* Berkeley: University of California Press.

Verluyten, S. Paul and Carina Hendrickx. 1987. La chute des consonnes finales facultatives en français: petite enquête sociolinguistique. *Linguistica Antverpiensia* **21**: 175-189.

Wendelken, Carter and Lokendra Shastri. ms. Probabilistic Inference and Learning in a Connectionist Causal Network.

Department of Linguistics
1203 Dwinelle Hall, UC Berkeley
Berkeley, CA 94720

bbergen@socrates.berkeley.edu