

Independence claims in linguistics

JOHN C. PAOLILLO
Indiana University

ABSTRACT

Empirical work in linguistics often puts forward claims about the independence of two phenomena as substantive hypotheses. But, independence is always an assumption in the framework of empirical hypothesis testing, meaning independence claims are not empirically verifiable. Hence, they need to be regarded differently from truly empirical hypotheses within the substantive theories of which they are part. In this paper, a number of independence claims are illustrated, alongside their problematic consequences for empirically guided theorizing. Recommendations are made regarding the use of independence that should facilitate the empirical testing of substantive linguistic hypotheses.

Much of linguistics is practiced as an amalgam of structuralist and rationalist practices, in which observation is regarded as a relatively uncomplicated process and emphasis is placed on deduction in the development of theoretical arguments. While clarity and soundness in rational argumentation are important in theory building, clarity in observation is also important. Used in conjunction, observation and reasoning constrain each other. But because of the greater emphasis on rationalism in linguistics, it is quite common for observation to be sacrificed in favor of some theoretical goal (e.g., parsimony, elegance). Consequently, rigor and verifiability are often compromised.

Variationist linguistics is a mode of empirical linguistic inquiry with more developed practices of observation than some other branches. Specifically, linguistic observations are not taken as solitary facts, usable in isolation; instead, they are regarded as potentially varying indications of more general tendencies. Statistical models of varying sophistication are used to sift apart the influences on any given observation that are either general or due to chance, and theorization directly addresses the tendencies that are generalized from the pooled observations. Variationist linguistics is thus an empirical social science. It uses the same statistical models and observational protocols as other empirical social sciences.

Nonetheless, the coupling of observation and theory is not always complete. One way in which a disconnect between observation and theory shows up is in the way that independence, in the statistical sense of the term, emerges as a central claim of a linguistic theory. Statistical independence is the default assumption in hypothesis testing. In fact, it is a basic axiom guiding inference in hypothesis testing. Hence, one does not empirically discover independence in the same way that one discovers nonindependence, and a reported finding of independence cannot be interpreted in the same way, either.

Independence claims come in a variety of forms in linguistics. Linguists often assume that social and linguistic environments in a variable process do not interact (Fasold, 1991), or that the rate of linguistic change is uniform (Kroch, 1989), or that the natural way to combine implicational hierarchies in Optimality Theory (Prince & Smolensky, 1993) is to “harmonically align” the two scales. All of these represent independence claims; all are potentially problematic in that they assert empirically untestable claims as if they were substantive, empirical hypotheses when they are not, and theory building around such concepts should take this into account.

The remainder of this paper is organized as follows. In the next section, statistical hypothesis testing is explained. Emphasis is placed on the logic of the significance test within a statistical modeling framework, and the place of statistical independence in relation to hypothesis testing. The third section outlines four types of independence claims in linguistics: (i) the noninteraction of social and grammatical constraints in variable rules, (ii) the Constant Rate Effect, (iii) implicational relations in two-way tables, and (iv) Harmonic Alignment in Optimality Theory. The final section summarizes the points presented about independence claims in linguistics and suggests some desiderata for theory building so that researchers can avoid asserting independence claims as empirical hypotheses.

HYPOTHESIS TESTING

Independence is an important concept in statistics, and it is fundamental to the construction and use of statistical models. Two events are said to be independent if knowing the outcome of one provides no information about the likely outcome of the other. Statistically, the probability of one event does not depend on the other; the conditional probability of the second event is the same, whatever the outcome of the first event. Statistical independence implies causal independence; hence, independence is a basic assumption, and any hypothesized dependence of events requires special evidence. The procedure for establishing this is known as a *hypothesis test*.

A hypothesis test applies probability theory as a guide to developing inferences, based on the distribution of some set of observations; it has a number of components: (i) a mathematical model of the expected distribution of observations, (ii) a stochastic component representing chance occurrence, and (iii) a test statistic whose chance distribution is known, measuring the fit of the model to the observed data. These components constrain each other and need to be chosen together in a way that is appropriate to the observations being made.

The mathematical model is typically composed of a number of terms involving *predictor variables*, or simply *predictors* (e.g., x_1 , x_2), corresponding to distinct aspects of each observation; these are known as “factors” in the variationist tradition and correspond to contextual conditions, whether social, linguistic, or otherwise. Each predictor variable is associated with a *parameter* (β) representing

the quantitative strength of the effect of the associated predictor. These are combined into a mathematical statement, for which the expression in (1) is typical.

$$f(y) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \varepsilon \quad (1)$$

In (1), y represents the variable of special interest, known as the *response variable*.¹ A typical response would be the relative frequency of t/d deletion in a sample of English. The function $f(y)$ is a *link function*, relating the predicted response variable to a sum of terms; because (1) has a link function, it is a Generalized Linear Model (GLM).² Logistic regression, commonly used in variationist analysis (Varbrul/GoldVarb), is a GLM where $f(y)$ is specified as $\ln \left[\frac{y}{1-y} \right]$ (McCullagh and Nelder, 1989; Paolillo, 2002).³ The term ε represents the *error distribution*, the model's stochastic component.

For any given situation, different model statements with different terms are possible, each making a distinct statement about the relations among the observations. For example, a simpler model than (1) could omit the $\beta_1 x_1$ term, the $\beta_2 x_2$ term, or both. Model statements omitting a term express the statistical independence of y and the omitted predictor x . Alternatively, a predictor x corresponding to any term whose β value is 0 is independent of y , because that mathematically accomplishes the same thing. Moreover, the terms of the model (1) are assumed to take their values independently of one another: x_1 and x_2 have independent effects on y in this model. There is no need to consult the value of x_1 to determine what the effect on y is for a given value of x_2 , or vice versa. Of course, a different model could include a term $\beta_{1,2} x_1 x_2$ whose value depends on the conjunction of the values of x_1 and x_2 ; such a term is called an *interaction term* and it represents an effect in which y , x_1 , and x_2 are all mutually dependent. The term $\beta_{1,2} x_1 x_2$ is nonetheless independent of the other model terms $\beta_1 x_1$ and $\beta_2 x_2$ in that there are distinct contributions to y for each of the three terms. Each term also corresponds to an independently testable hypothesis about the relationship between the x values and y .

Hypothesis testing builds on this kind of model comparison directly by designating some simpler model as a *null hypothesis* (H_0). Classically, H_0 is described as an absolutely minimal model with only a chance component (e.g., $f(y) = \varepsilon$, meaning the observed y is independent of all other observed events), but any reasonable model may have the status of H_0 for a specific test. Most commonly in a model such as (1), terms of the model are tested independently, and for each test, H_0 represents $\beta = 0$ for some specific term; that is, the test asks whether the term needs to be included in the model or not.

The hypothesis test is performed by computing a test statistic and comparing its value with its expected chance distribution, derived from ε .⁴ We then compare the test statistic to an established criterion value, with a sufficiently small probability. Values of the test statistic whose probability is less than the criterion value are considered to indicate that H_0 is inadequate and should be rejected; such values are called "significant." Values whose probability is greater than the chosen criterion value are regarded as nonsignificant, and the model is retained as

satisfactory. When H_0 represents $\beta = 0$, significance means that the term needs to be retained (y is dependent on x), and nonsignificance means that y can be treated as independent of the corresponding x . For interaction terms, the interpretation is similar: if H_0 ($\beta_{1,2} = 0$) is not rejected, y is assumed to be independent of the conjunction of x_1 and x_2 ; otherwise, y , x_1 , and x_2 are mutually dependent.

Key to statistical hypothesis testing is the notion of rejecting H_0 under evidence that it is unlikely to be true. There is an asymmetry between significant and nonsignificant results in this respect. For significant results, H_0 is rejected, leading the researcher to propose an alternative model. For nonsignificant results, H_0 is retained, but because it was assumed to begin with, its validity has not actually been empirically established by the significance test, but rests instead upon the reasonableness of any arguments that were used to suggest it in the first place. Furthermore, H_0 always represents some kind of independence: y is independent of some x , or y is independent of some conjunction of x 's. Hence, independence is always the default assumption in any hypothesis test.

Types of error

A crucial notion in significance testing is the notion of error, by which we mean an incorrect inference. Since statistical inference is based on a probability, specifically a probability associated with H_0 , it is guaranteed that some statistical inferences we make will be incorrect. Yet, because we are able to know the probability associated with these errors, it is possible to understand their risk and optimize it in the interest of soundness of inference, on the one hand, and attention to theoretically interesting phenomena, on the other. Normally, this optimum is selected by adjusting the criterion probability for a significance test.

To understand how this optimization works, we need to distinguish two types of incorrect inference, known as type 1 and type 2 errors. A type 1 error occurs when H_0 is falsely rejected in favor of an alternative hypothesis, when in fact the independence represented by H_0 holds. A type 2 error occurs when H_0 is not rejected, even though it does not hold, and some alternative hypothesis H_1 would be more appropriate. Type 1 and type 2 errors are complementary, and there is a tradeoff between them such that as the probability of type 1 error decreases, the probability of type 2 error increases.⁵ But, only the distribution under H_0 is actually known, because the hypothesis space for alternatives to H_0 is uncountably infinite and cannot be known. Hence, in hypothesis testing, we adjust the criterion level to be conservative with respect to the probability of type 1 error and are forced to be agnostic about the probability of type 2 error.⁶ In practice, criterion probabilities (often called p levels) of .05 or .01 are typically taken to be acceptable, with the latter being a stricter criterion requiring stronger evidence to reject H_0 .

The distinction between type 1 and type 2 errors allows us to make an important connection between different types of substantive hypotheses and the kinds of inferences we can draw from a hypothesis test. One type of hypothesis asserts that two phenomena are mutually dependent in some way; either one causes the

other, or the two arise together from some common cause. Such a hypothesis will be represented in a model by a term relating some predictor x to the observed y . The null hypothesis H_0 is represented by $\beta = 0$, which is readily tested by a statistical hypothesis test; the type 1 error, under assumed independence, is known. Such hypotheses, positing dependence, are therefore empirically testable.

A second type of hypothesis proceeds in the other direction, arguing that there is not *any* kind of statistical dependence between two events. This requires taking some unspecified H_1 as the starting point and arguing for independence. However, it is not clear what H_1 to start with (there are uncountably many), and so we cannot know the probability distribution of the test statistic under H_1 . Actually, we must consider all possible H_1 , and we must show all of them to be unlikely; many have effects that are immeasurably small, and so not empirically distinct from H_0 . In other words, we cannot know the probability distribution of type 2 error, the way we can for type 1, and we cannot set a criterion probability for it. Hence, claims of independence are hypotheses that are not empirically testable.

The role of independence in a hypothesis makes an important distinction between two types of hypothesis and the roles they can have in theory building, when empirical procedures are used to guide the process. Claims of dependence are empirically testable, and contrast with a default position of assumed independence. Claims of independence are not empirically testable. When they arise in a theory, they arise as assumptions and are not guided by empirical principles.

CLAIMS OF INDEPENDENCE

We turn now from the statistical view of independence to considering specific cases in which statements of independence are made as substantive theoretical claims. Independence claims take different forms and are not always easy to recognize. Moreover, some forms of independence claims actually appear to be something else, like logical implication. Other times, independence claims are embedded within other claims, and some effort is required to disentangle them. Nonetheless, independence claims are common in linguistics. In this section, we examine four different claims of independence and illustrate the problems they raise for linguistic theorizing. We begin with the most transparent types of independence claims, and proceed toward forms of independence claims that are less readily recognizable.

Noninteraction of internal and external constraints

An oft-repeated claim in the variationist literature is that internal (linguistic) and external (social) constraints on variation are independent. The claim is found in both substantive and methodological versions, and their consequences are somewhat different. The ultimate source of these claims appears to be Labov; one version appears in Fasold (1991), citing Labov (no reference given), and a

nearly identical formulation is given in Labov (1994) as justification for the organization of *Principles of Linguistic Change*, in which “internal factors” and “external factors” are each explored in separate volumes published seven years apart.

The multivariate analyses that are included in these volumes almost always include both internal and external factors. When the analyses are carried out, it appears that the two sets of factors—internal and external—are effectively independent of each other. ... Moreover the internal factors are normally independent of each other, while the external factors are heavily interactive. These basic sociolinguistic findings provide the methodological rationale for the way the material in volumes 1 and 2 is divided and for the separate discussion of internal and external factors. (Labov, 1994:3)⁷

On the substantive level, Labov wants to be able to discuss types of linguistic changes (chain shifts, mergers, and splits) separate from their contexts, so that questions posed by structural and historical linguists can be addressed. On the methodological level, treating social and linguistic constraints separately has advantages for the multivariate models employed, because different predictor variables are more readily treated as independent, and interactions or variable nesting patterns raise problems of combinatoric complexity.

The claim has problems on both the empirical and analytical levels, however. On the empirical level, the claims “social and linguistic factors do not interact” and “linguistic factors do not interact” are both independence claims; they are assumed by default as long as there is no evidence to reject them. Accepting them is subject to a type 2 error of unknown probability, is therefore never empirically supported, and is only as reasonable as the supporting arguments. Sigley (2003) evaluated these claims for t/d deletion in a specific corpus of New Zealand English and demonstrated several social-by-linguistic and linguistic-by-linguistic factor interactions. Hence, Sigley argued, it is unsafe to assume noninteraction in general, and for that reason, he recommended a regime of systematic and thorough testing of interaction effects.

On the analytic level, however, there is a deeper problem, which amplifies Sigley’s criticism. At some point prior to the multivariate analysis, there is always an identification of the object of study, that being a variable linguistic process. This identification places both linguistic and social bounds on the problem, framing the data collection and subsequent analysis. The effect of this definition is to partition consideration of potentially related phenomena into different analyses. This procedure provides no guarantee that, for example, lack of observed interaction, is anything other than a methodological artifact. Moreover, otherwise observable interactions are potentially obscured by being presented in separate analyses.

For example, consonant cluster reduction in English affects specific segments (t/d) in specific linguistic contexts (e.g., following consonant, vowel, or pause), which were studied by Guy (1980) in New York City and Philadelphia. The two

different speech communities (social factors) show a common variable process, except for a difference in the treatment of following pause: in Philadelphia, it favors retention, whereas in New York City, it favors deletion. Had the two analyses been combined into a common multivariate analysis, this result would be represented by a significant social-by-linguistic interaction for following pause within dialect group. In other words, the observation of different factor weights across examples of related phenomena in different speech communities, where the differences are meaningful, corresponds to the identification of significant interactions between social and linguistic factors in a variable process.

A similar criticism applies to the failure to observe interactions among internal constraints. As before, independence is the default assumption, to which the analyst returns, when evidence for interaction has not been found, and claiming noninteraction is subject to an unknown probability of type 2 error. Moreover, linguistic processes are identified in ways that partition them from other, potentially related processes; to the extent that this involves the identification of contrasting effects in distinct but related processes (e.g., lenition affecting one class of consonants, but not another, even in the same environment), this implies the identification of significant nonindependence relations among linguistic predictors.

This criticism is general, applying equally to all linguistic variables and changes in progress. In every case, there is a linguistic variable of interest, which occurs in specific linguistic environments. The process is observed in a specific speech community, or contrasts linguistically across communities or structural segments of the same community. Hence, identification of interesting instances of nonindependence of internal and external factors (or among internal factors) is central to the practice of variationist analysis. Assumed independence of internal and external factors may be methodologically helpful, but only by introducing a methodological artifact, and substantively, the claim is vacuous.

Constant rate effect

A second independence claim is made by Kroch (1989 and related work) called the constant rate effect (CRE). The CRE basically says that when a linguistic change occurs, it propagates through all relevant linguistic environments at the same rate. Various examples are presented to illustrate this effect; in all cases, the argument is made that the regression slopes of the change over time are the same in all linguistic environments, and that, therefore, the change proceeds at a constant rate, irrespective of the environment in which it is examined. The model used is a logistic regression model such as (1). A recent reinterpretation of the CRE is framed in terms of constant *curvature* (Kallel, 2007). By using a more sophisticated statistical model, the rate of change over time is allowed to vary; although at any given point in time, it is the same in all environments. This reinterpretation is also an independence claim, subject to the same problems as the CRE.

As an example of the CRE, consider the case of *do*-periphrasis in English, for which Kroch (1989) reanalyzed data originally produced by Ellegård (1953),

TABLE 1. *Frequency of do-periphrasis in English by time period*

Time period	Negative declaratives		Negative questions		Affirmative transitive questions		Affirmative intransitive questions		Affirmative WH questions	
	%	Total	%	Total	%	Total	%	Total	%	Total
1400–1425	0	177	11.7	17	0	3	0	7	0	1
1425–1475	1.2	903	8.0	25	10.7	56	0	86	0	27
1475–1500	4.8	693	11.1	27	13.5	74	0	68	2.0	51
1500–1525	7.8	605	59.0	78	24.2	91	21.1	90	11.3	62
1525–1535	13.7	651	60.7	56	69.2	26	19.7	76	9.5	63
1535–1550	27.9	735	75.0	84	61.5	91	31.9	116	11.0	73
1550–1575	38.0	313	85.4	48	73.7	57	42.3	71	36.0	75
1575–1600	23.8	629	64.8	128	79.2	173	44.4	205	38.3	120
1600–1625	36.7	278	93.7	95	77.3	277	61.9	310	29.8	171
1625–1650	31.7	344	84.2	38	90.9	66	75.7	74	53.0	66
1650–1700	46.0	274	92.3	52	94.7	76	70.2	131	54.9	51

Source: Reproduced, with permission, from Kroch (1989:Table 3), who used data from Ellegård (1953).

which is reproduced in Table 1. The data were submitted to a logistic regression analysis, from which no difference was found in the rate of increase of *do*-periphrasis over time in the five environments. Each environment has a different frequency of periphrasis at any given time, and that effect is constant and entirely independent of the rate at which *do*-periphrasis is changing. That the CRE is an independence claim is very clearly stated in Kroch (1989).

Thus, if a study reports a series of multivariate analyses for different time periods, and the contextual effects are constant across these analyses, the rate of change of each context measured separately would necessarily be the same. This equivalence holds because, in statistical terms, the constant rate hypothesis is the claim that the overall rate of use of a form is independent of the contextual effects of its use. (Kroch, 1989:204)

Kroch then refers to the functional form of the logistic regression model (1) in which the independent contextual factors present in a context are represented by distinct weights added together, concluding, “As is clear from the equation, the contextual effects are constant across time and do not interact with the time variable” (Kroch, 1989:204). In other words, the fact that there is no $\beta_{context,time}x_{context}x_{time}$ interaction term in the model means that change over time is constant in all contextual environments. Note that the appropriate null hypothesis in this test is that the regression slopes for individual contexts are the same: in other words, *we assume that the rate of change is constant*, and only reject this assumption if we find evidence otherwise. Failure to find evidence to refute this assumption is not unequivocal support of it. Furthermore, by retreating to the null hypothesis, we invite type 2 error of unknown probability. The strength of the CRE, as a substantive hypothesis, therefore, rests solely on

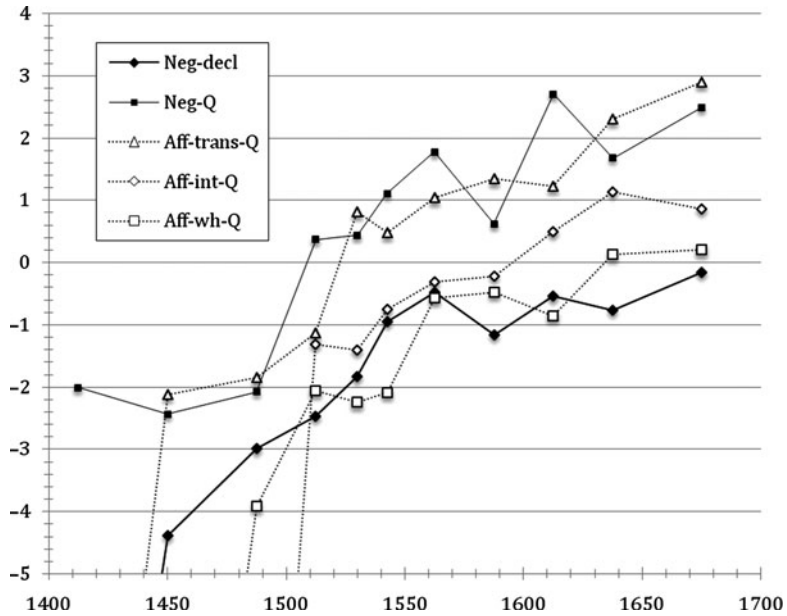


FIGURE 1. Frequency of *do*-periphrasis in English, measured in empirical logits. *Source*: Data from Ellegård (1953), as reported in Kroch (1989:Table 3).

its reasonableness within the substantive domain of linguistics, and not on the quantitative observation of independence.

To see more clearly how the argument for the CRE needs to be made, consider Figure 1, in which the frequency of *do*-periphrasis from Table 1 is represented on the logit scale, and each time period is plotted at the center of its corresponding time interval. This is the scale on which the logistic regression is estimated, and constancy in the rate of change would be observed as parallel regression lines in this scale. Note that the empirical logits of cells in Table 1 with 0% periphrasis are $-\infty$, and are represented here by the relevant lines trailing off the graph.

For each series in Figure 1, a regression slope needs to be estimated, but there is considerable variation in all of them, such that they cross each other in multiple places, and it is not entirely clear where the “true” lines should fall or what their slopes should be. These slopes are likely to be very similar, and given the variation observed and the small *N*s for some of the series, it is highly likely that their confidence intervals overlap. Consequently, a significance test is unlikely to establish the need for individual time-context regression slopes. In other words, we accept the null hypothesis that regression on the five individual series has the same slope as regression on the series aggregated together.

The principal cause of the failure to find nonindependence in the data is the variance of the different series, whose wandering easily overwhelms any estimate of difference between their slopes. Historical data are never randomly sampled, and sampling inadequacies (and potentially biases) very probably

confound Ellegård's data. Kroch (1989) did not discuss this aspect of the *do*-periphrasis data, in spite of its relevance to the observed variability and the apparent inability to observe different rates of change in different contexts. This was an oversight, as consideration of sampling deficiencies should be prior to consideration of any substantive hypotheses such as general principles of language change.

As regards independence and the CRE, it is immaterial how many times we fail to make an observation of differential rate of change. Constant rate of change is assumed, rather than established by empirical finding, and claiming independence invites type 2 error of unknown probability. Because different observational inadequacies apply to different studies of the CRE, no general argument can be made regarding the adequacy of sampling, among other things, for all cases in which the CRE is observed. Finally, as with the independence of internal and external factors in language change, the analyst has methodological freedom to operationalize the object of study in ways that partition consideration of related changes into different historical processes. At a very minimum, there is something that changes in the system (e.g., *do*-periphrasis) and something that does not (parts of the verbal system not involving *do*). The availability of such analytical options means that the analyst can always posit an analysis consistent with the CRE. Hence, the CRE is an assumption, not an empirical hypothesis, and it cannot be meaningfully tested using empirical means.

Implicational relations

Another type of circumstance in which linguists make independence claims is in places where an implicational relation is asserted. The involvement of statistical independence is less obvious than in other types of claims, because in these cases, the outcomes involved are typically discrete, and the theories that are proposed to account for them often do not admit variable outcomes. Categorical variation of this type does not immediately appear to be consistent with variationist analysis, and analysis is cast using logical relations instead, such as implication. However, there is actually no inconsistency, and different types of circumstances in categorical analyses correspond to different statistical model statements. Some statements using logical relations turn out to be independence claims when examined this way and, hence, are less interesting than they are commonly thought to be.

The basis for the statistical treatment of logical relations explained here comes from works by Sankoff and Rousseau (e.g., Rousseau, 1989; Rousseau & Sankoff, 1978a, 1978b; Sankoff & Rousseau, 1974, 1980, 1981) that treated implicational scales within the variable rule framework, using the logistic regression model as given in (1). Their basic observation can be stated as follows. Categorical variation is a special case of variation in which the β parameters take extreme values, such that the nonlinear logistic transformation results in categorical predictions (probabilities of 1 or 0). For the logistic transform,⁸ values between 7 and 8 on the logit scale converge to 1 on the probability scale (within three

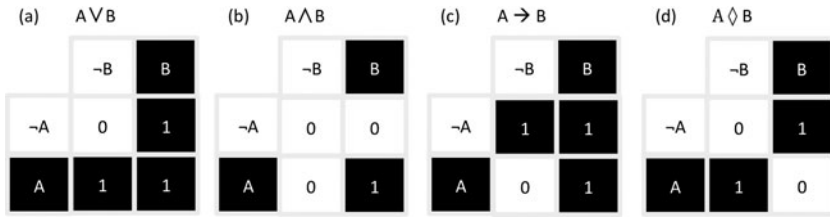


FIGURE 2. Logical relations in two-way tables.

TABLE 2. Logistic regression models for different logical connectives

Fig. 2 panel	Formula	Model Parameters				Cells			
		α	β_A	β_B	$\beta_{A,B}$	$\neg A, \neg B$ (α)	$\neg A, B$ ($\alpha + B$)	$A, \neg B$ ($\alpha + A$)	A, B ($\alpha + A + B + A, B$)
(a)	$A \vee B$	-8	+16	+16	0	-8	+8	+8	+24
(b)	$A \wedge B$	-24	+16	+16	0	-24	-8	-8	+8
(c)	$A \rightarrow B$	+8	-16	+16	0	+8	-24	+8	+8
(d)	$A \diamond B$	-8	+16	+16	-32	-8	+8	+8	-8

decimal places), meaning that for any combination of factors where the βx terms combine to yield a value greater than 7 (or less than -7), more than 1000 observations would likely be needed in that specific context to observe a variable outcome.

The relation between logical implication and independence can be seen in Figure 2. Each panel represents a two-by-two table of relative frequencies in which cells with a value of 0 are those in which the response is categorically absent, whereas those with a value of 1 have the response categorically present. The rows and columns represent the absence and presence of two contextual predictors, labeled A and B for generality. The four panels have patterns corresponding to the truth tables of four logical connectives: disjunction ($A \vee B$), conjunction ($A \wedge B$), implication ($A \rightarrow B$), and exclusive disjunction ($A \diamond B$). These patterns are presented as parameterizations of the logistic regression model in (1) in Table 2, in which the center four columns give the parameters of each model, and the last four columns give the cell predictions for each, on the logit scale. Cell values exceeding +7 or -7 yield categorical predictions on the proportion scale (1 or 0), when rounded to three decimal places.⁹ Because all the cells do, only the signs (+/-) matter when reading the predictions on the proportion scale, with positive (+) values corresponding to 1s and negative (-) values corresponding to 0s.

Each of the first three rows of Table 2 represents a model with no interaction term, because $\beta_{A,B}$ in each is 0; predictors A and B represent independent contextual effects in these models. Only the last model has a non-0 interaction

term, so *A* and *B* are nonindependent in it. Notably, model (c), which corresponds to logical implication, requires no interaction parameter to state in the logistic regression model.

Consider a claim to the effect that the existence of a variant (1) in a context with a predictor *A* implies the existence of a predictor *B* in the context, and *A* without *B* categorically shows the alternative (0). Predictors *A* and *B* are general; they could be speakers, social strata, grammatical environments, or any other relevant predictor that is part of a grammatical model. The only important fact about them is their implicational relation. In just such a case, the relation among *A*, *B*, and the variant is represented exactly by panel (c) in Figure 2, with the associated model parameters in the third row of Table 2 expressing the same relation. In other words, an implicational relationship between two predictors is expressed as a model in which *A* and *B* have (only) independent effects, and therefore, a claim to the effect that *A* implies *B* is an independence claim.

As a concrete example of this, consider the comparison of subject selection in Chamorro (Chung, 1998) and Lummi (Jelinek & Demers, 1983) as analyzed by Aissen and Bresnan (2002b). The relevant variation in both cases is active and passive voice; Aissen and Bresnan discussed this in terms of “subject selection” (passive and active select a particular argument as subject of the sentence), governed by constraints framed within optimality theory (OT) (Aissen 1997; Aissen & Bresnan, 2002a; Bresnan, Dingare, & Manning, 2001). In Chamorro, the grammatical conditions on active and passive voice are animacy and thematic role, whereas in Lummi, they are proximity (grammatical person deixis) and thematic role. The animacy (animate/inanimate) and proximity (first or second versus third person) of both agent and patient thematic roles are independent of each other, allowing four relevant combinations in each language, as indicated in Table 3. Each of these combinations may realize active, passive, or both, in which case there is variation.

Taking passive as the 1 variant and active as 0, and provisionally treating cells in which active and passive occur together as .5 (0 on the logit scale), logistic regression models can be constructed for the observed variation in both languages as indicated in the lower half of Table 3. Note that both languages exhibit independence of agent and patient (with respect to animacy or proximity), as there is no interaction parameter ($\beta_{a,agt,b,pat} = 0$, for any *a* and *b*). Furthermore, any intermediate level of variation in the variable cells can be obtained by adjusting the parameters. For example, if the proximal agent, proximal patient cell in the upper left of the Lummi table had 40% active voice and 60% passive, $\beta_{prox,agt}$ could be changed to -8.405 ; with all other parameters the same, the predicted relative frequency would change only in that specific cell. Both the Lummi and Chamorro examples illustrate independence, and the implicational relations observed in Table 3 are independence claims, representing the null hypothesis that $\beta_{a,agt,b,pat} = 0$. As independence claims, they can only be rejected, not verified, by empirical data, and acceptance of the null hypothesis has an unknown probability of type 2 error. The only content present

TABLE 3. *Subject selection (voice) and animacy or proximity in Chamorro and Lummi*

	Chamorro			Lummi	
	Animate patient	Inanimate patient		Patient 1st or 2nd	Patient 3rd
Animate Agent	Active and passive	Active only	Agent 1st or 2nd	Active only	Active only
Inanimate Agent	Passive only	Active and passive	Agent 3rd	Passive only	Active and passive
Parameters	α	0	Parameters	α	0
	$\beta_{ani.agt}$	-8		$\beta_{prox.agt}$	-16
	$\beta_{ani.pat}$	+8		$\beta_{prox.pat}$	+8
	$\beta_{ani.agt,ani.pat}$	0		$\beta_{prox.agt,prox.pat}$	0

Source: Adapted from Aissen and Bresnan (2002b).

TABLE 4. *Generating constraint orders and parameter weights under HA: Original preference orders are $a > b > \dots > z$ and $X > Y$*

Preference scales	A	>	b	>	...	>	z	
X	X/a	>	X/b	>	...	>	X/z	Harmonic scales
Y	Y/a	<	Y/b	<	...	<	Y/z	
Constraint orders	*X/a	«	*X/b	«	...	«	*X/z	
	*Y/a	»	*Y/b	»	...	»	*Y/z	
Candidates	A		b		...		z	
X	$\beta_{X/a}$	>	$\beta_{X/b}$	>	...	>	$\beta_{X/z}$	
Y	$\beta_{Y/a}$	<	$\beta_{Y/b}$	<	...	<	$\beta_{Y/z}$	
Parameters	β_a	>	β_b	>	...	>	β_z	
α_X	A		b		...		z	

in the implicational relations of Table 3 resides in the relative magnitude of the agent and patient effects on subject selection.

Harmonic alignment

Aissen and Bresnan (2002b) did not merely cast the observation of passive/active voice in Chamorro and Lummi in terms of implicational relations in two-by-two tables. Rather, they sought to motivate these observations in terms of a principle of OT known as Harmonic Alignment (HA) (Aissen 2003; Bresnan, et al., 2001). But Rousseau and Sankoff’s (1978b) treatment of categorical variation also generalizes to larger two-way and multiple-way classifications, including those implied by OT (Paolillo, 2002, 2010). Categorical variation in such systems is therefore analyzable in terms of the logistic regression model in (1), and the same potential for independence claims arises in the general case. This can be seen by closer examination of the principle of HA.

Harmonic Alignment is appealed to in various applications of OT, both in syntax (e.g., Aissen, 2003; Aissen & Bresnan, 2002a) and phonology (e.g., Anttila, 1997). Harmonic Alignment proposes that a pair of preference scales combine more naturally (“harmonically”) with the preferred items on both scales aligned, as worked out by generating “harmony scales” that are subsequently used to generate constraint rankings. This process is defined when one or both of the original preference scales is binary.¹⁰ We begin with two preference scales $X > Y$ and $a > b > \dots > z$. These combine to produce two harmony scales: $X/a \phi X/b \phi \dots \phi X/z$, having the preferred element of A combined with each of the elements in second scale in the original order, and $Y/z \phi \dots \phi Y/b \phi Y/a$, with the dispreferred element of the first scale and the elements of the second scale in reverse order. These two harmony scales are interpreted as specifying an ordering of (negative) constraints with the order of the rankings of both scales reversed once more: $*X/z \phi \dots \phi *X/b \phi *X/a$ and $*Y/a \phi *Y/b \phi \dots \phi *Y/z$. The central claim of HA is that these orderings of negative constraints, as derived from the original preference rankings, are fixed. The two sets of constraint rankings, however, are not ordered with respect to each other, and they may be interleaved to get different constraint rankings for different analyses. Anttila (1997) combined this with “crucial unranking” to predict variation in the Finnish genitive plural.

To cast these statements in terms of a statistical model, we first recognize that conjoined constraints of this type essentially name each of the cells independently as interaction effects: $*X/a$ corresponds to the conjunction of X and a in a specific (candidate) form. Constraint ranking simply requires the assigned weights to be in a specific order, leaving their magnitudes (the exact β values) unspecified. We can ask, therefore, if a model with independent parameters will fit this specification of cell values. The relevant model is a logistic regression model in which the binomial value X or Y is modeled with respect to the predictor a, b, \dots, z ; the relevant parameters are found by taking the ratio of the relevant $\beta_{X/}$ and $\beta_{Y/}$ values, such as $\beta_a = \beta_{X/a} \div \beta_{Y/a}$. Because the orders of the values $\beta_{X/a} > \beta_{X/b} > \dots > \beta_{X/z}$ and $\beta_{Y/a} < \beta_{Y/b} < \dots < \beta_{Y/z}$ are fixed (in opposite directions), the corresponding ratios $\beta_a > \beta_b > \dots > \beta_z$ are also fixed in order, reflecting the original preference order $a > b > \dots > z$. In a logistic regression model, one of these usually takes the value 0, and a parameter α_X describes an overall reference value for the rate of X vs. Y preference. Harmonic Alignment, therefore, derives, through a circuitous process, a constraint order equivalent to a logistic regression model with an order of parameter values identical to the original preference orders.

Hence, Harmonic Alignment adds no information beyond the original preference orders and represents an independence claim, a null hypothesis about the relation between constraint rankings. It is not empirically verifiable, and acceptance of Harmonic Alignment is subject to an unknown probability of type 2 error. As with the CRE and other independence claims, HA imposes no particular requirement on the way that preference orders or the linguistic predictors so ordered are composed. Consequently, if potentially falsifying data were discovered, HA is readily maintained by redefining the linguistic predictors

or constraints.¹¹ Therefore, HA should not be regarded as a statement that makes substantive empirical claims about language.

INDEPENDENCE AND THEORY CONSTRUCTION

We have seen four examples of independence claims from linguistics, noting that they are problematic from an empirical perspective, because they are not testable and may only be assumed true. Other assumptions consistent with the observations are possible and, hence, may turn out to be true instead, leading to an unknown probability of type 2 inference error. The reasonableness of an independence claim for whatever linguistic phenomenon being considered depends solely on things other than the observations being made.

In some cases, the independence claims are easily refuted. In others, the independence assumptions are buried under analytical assumptions or metalanguage that are not transparently mapped onto statistical concepts. Consequently, independence claims may arise in ways that are difficult to appreciate or to guard against. How then does one address the role of independence in theorizing so that these problems do not occur? The foregoing examples do not exhaust the places in linguistic theorizing where independence claims can arise. More examples can be cited, each of which appears reasonable, and many have stalwart adherents who would need to be convinced of the nonempirical status of theirs and others' independence claims. Independence claims are therefore not unmistakably obvious when they occur. However, it is possible to identify certain "danger zones" where independence claims are likely to occur.

One such danger zone involves theoretical constructs enumerating combinatorial possibilities for some kind of phenomenon. Harmonic Alignment is one such example, but OT offers yet others. The function GEN, which generates a set of candidates for an input form is one, and the permutation of partial constraint rankings (as generated under HA) into more fully specified orders or "factorial typologies" is another (see, e.g., Fong & Anttila, 2010). The formal devices used to explore such combinatorial systems (Cartesian products, Galois lattices) may be used to similarly illuminate statistical models,¹² and a careful mapping across these different abstract domains can often profitably illuminate the relationship between a theoretical model and the family of statistical models it corresponds to. Therefore, one can regard entities such as GEN as formal devices that permit one to work out the consequences of an analysis. At the same time, linguists have a tendency to overinterpret them, so that independence becomes a theoretical claim rather than a guide to finding interesting substantive relationships in data. For this reason, combinatoric generators need to be regarded with some care, so that inappropriate independence claims are not derived from them.

A second way in which independence becomes a claim arises when empirical observations are made and a null hypothesis is not refuted in some presumably

large number of cases. Examples of this type are the noninteraction of social and linguistic factors, and the Constant Rate Effect. Independence is inherently present in the statistical models used as the null hypothesis regarding the interrelation of multiple factors. Because refuting the null hypothesis requires data, and datasets used in studies of linguistic variation are small and noisy, it can be quite difficult to refute the null hypothesis. Alternatively, redefining the object of inquiry masks observed dependencies. At some point, a researcher may suspect that the null hypothesis will never be refuted, leading to an interpretation of the assumed model of independence as a substantive result of the empirical analysis, when it is in fact not a substantive result and can never be one. Disconfirmation of an independence claim does lead to a Popperian kind of progress in inquiry, but a failure to disconfirm merely leaves one with one's starting assumptions, and given that the reasons for such failure are many (e.g., insufficient data, poor research design), little can really be concluded from it.

How then does one use independence in empirical reasoning? Independence is a central characteristic of statistical models, where it expresses the notion that different factors can have distinct, simultaneous contributions to an observed phenomenon. This notion corresponds to a relatively simple understanding of causality, which is certainly simpler than one in which complex conditions are required to decide which causes are relevant to a phenomenon before making a prediction. The latter possibility corresponds to some form of interaction, whose representation is more complex, and which we want to have persuasive evidence for before positing. In other words, independence represents our starting point when reasoning about multiple potential causes for a phenomenon. As a null hypothesis, it can guide us into making discoveries when evidence permits us to falsify it, but it makes no theoretical contribution of which we should be overly credulous. Therefore, it is an error to incorporate independence claims into the architecture of an explanatory theory or to use them in any way where they are subject to overinterpretation.

The arguments presented here stress that the notion of statistical independence needs to be properly understood with respect to its role in empirical research. Independence claims commonly arise from the combinatorics of our conceptual models and from observations that fail to refute a null hypothesis. Claims to the effect that two phenomena are necessarily independent generally represent a failure to recognize these circumstances where they occur. Independence may have other useful roles in theorizing, such as in the deductive processes that map out the consequences of a formal model, but in and of themselves independence claims make little useful contribution to the empirical side of the research.

NOTES

1. Another term is *dependent variable*, contrasted with *independent variable*, for which we use the term *predictor* here. Although there is a tradition behind these terms, they confound discussion of statistical independence, so we do not use them here. Other statistical traditions use the term *feature*.
2. Link functions in GLMs are restricted to members of the exponential family; see McCullagh and Nelder (1989) for details.
3. Other models like those of factor analysis have different functional forms and are not discussed here for reasons of space.

4. In a GLM, two such tests are available: the analysis of variance test, and the Wald test. In analysis of variance, the test statistic is computed by aggregating the deviances of the observed y from the estimated \hat{y} , and in turn referencing the aggregate to either the F -distribution (for linear models) or the chi-square distribution (for logistic regression and log-linear models). The Wald test is computed from the β values directly and uses the normal or t -distributions.
5. A similar trade-off occurs in decision theory in psychology, when a decision criterion is adjusted (Swets, 1998), as well as in signal detection theory, both of which differ from statistical hypothesis testing in assuming a forced choice between exactly two alternatives whose probability distributions are both known. In statistical hypothesis testing, only the distribution under H_0 is known.
6. In the extreme case, if we set the type 1 error probability to 0, we will never reject H_0 and, therefore, will be forced to make at least some type 2 errors. Other factors go into the choice of p , including the number of tests to be run (see Sigley, 2003).
7. Weinreich, Labov, and Herzog (1968) make numerous statements that contrast with this assumption, although methodological versions of the independence assumption can be found there as well.
8. The functional inverse of the logit is the logistic, defined as $logistic(x) = e^x / (1 + e^x)$.
9. Sankoff and Rousseau's general proof involves a limit as the parameter weights go to infinity; for purposes of illustration, small finite values are used here.
10. Aissen (2003) develops a version of Harmonic Alignment that works with two nonbinary preference scales. The resulting hierarchy is a two-way table (2003:Figure 4) and its analysis is readily seen to be an implicational scale/model of independence of exactly the form described in Rousseau and Sankoff (1978b).
11. Note that the predictors in the Chamorro and Lummi examples are actually conjunctions (e.g., agent and animate, patient and animate), so a process of definition has already occurred, arriving at complex predictors.
12. A statistical model with only independent factors is a member of a join lattice with a single join operator, whereas a statistical model with interactions is a member of a join lattice with two distinct join operators.

REFERENCES

- Aissen, Judith. (1997). On the syntax of obviation. *Language* 73(4):705–750.
- . (2003). Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory* 21:435–483.
- Aissen, Judith, & Bresnan, Joan. (2002a). Optimality and functionality: Objections and refutations. *Natural Language and Linguistic Theory* 20:81–95.
- . (2002b). Harmonic alignment in morphosyntax: Subject selection. Course handout from Special Joint Summer School (Linguistic Society of America/Deutsche Gesellschaft für Sprachwissenschaft), Düsseldorf, Germany. 2002. Available at: <http://www.phil-fak.uni-duesseldorf.de/summerschool2002/LNAissen1.pdf>.
- Anttila, Arto. (1997). Deriving variation from grammar. In F. Hinskens, R. van Hout, & L. Wetzels (eds.), *Variation, change and phonological theory*. Amsterdam: Benjamins. 35–68.
- Bresnan, Joan, Dingare, Shipra, & Manning, Christopher D. (2001). Soft constraints mirror hard constraints: Voice and person in English and Lummi. In M. Butt & T. H. King (eds.), *Proceedings of the LFG 01 Conference*, 13–32. Stanford: CSLI Publications. <http://csli-publications.stanford.edu/LFG/6/lfg01.html>
- Chung, Sandra. (1998). *The design of agreement: Evidence from Chamorro*. Chicago: University of Chicago Press.
- Ellegård, Alvar. (1953). The Auxiliary do: The establishment and regulation of its use in English. In F. Behre (ed.), *Gothenburg studies in English*. Stockholm: Almqvist and Wiksell.
- Fasold, Ralph. (1991). The quiet demise of variable rules. *American Speech* 66(1):3–21.
- Fong, Vivian, & Anttila, Arto. (2010). Variation and ambiguity. In L. Uyechi & L.-H. Wee (eds.), *Reality exploration and discovery: Pattern interaction in language and life*. Stanford, CA: CSLI Publications. 345–358.
- Guy, Gregory. (1980). Variation in the group and in the individual. In W. Labov (ed.), *Language variation in space and time*. New York: Academic Press. 1–36.
- Jelinek, Eloise, & Richard A. Demers. (1983). The agent hierarchy and voice in some Coast Salish languages. *International Journal of American Linguistics* 49:167–185.
- Kallel, Amel. (2007). The loss of negative concord in Standard English: Internal factors. *Language Variation and Change* 19(1):27–49.

- Kroch, Anthony. (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change* 1(3):199–244.
- Labov, William. (1994). *Principles of linguistic change*. Vol. 1. *Internal factors*. Oxford: Blackwell.
- McCullagh, Peter, & Nelder, John A. (1989). *Generalized linear models*. 2nd ed. Boca Raton, FL: Chapman & Hall/CRC.
- Paolillo, John. (2002). *Analyzing linguistic variation: Statistical models and methods*. Stanford, CA: CSLI Publications.
- . (2010). Optimality theory as a probabilistic model. In L. Uyechi & L.-H. Wee (eds.), *Reality exploration and discovery: Pattern interaction in language and life*. Stanford, CA: CSLI Publications. 105–124.
- Prince, Alan, & Smolensky, Paul. (1993). *Optimality theory: Constraint interaction in generative grammar*. Piscataway, NJ: Rutgers University Center for Cognitive Science.
- Rousseau, Pascale. (1989). A versatile program for the analysis of sociolinguistic data. In R. Fasold & D. Schiffrin (eds.), *Language change and variation*. Amsterdam: Benjamins. 395–409.
- Rousseau, Pascale, & Sankoff, David. (1978a). Advances in variable rule methodology. In D. Sankoff (ed.), *Linguistic variation: Models and methods*. New York: Academic Press. 57–69.
- . (1978b). Singularities in the analysis of binomial data. *Biometrika* 65(3):603–608.
- Sankoff, David, & Rousseau, Pascale. (1974). A method for assessing variable rule and implicational scale analyses of linguistic variation. In J. Mitchell (ed.), *Computers in the humanities*. Edinburgh: Edinburgh University Press. 3–15.
- . (1980). Categorical contexts and variable rules. In S. Jakobson (ed.), *Papers from the Symposium on Scandinavian Syntactic Variation*. Stockholm: Almqvist and Wiskell International. 7–22.
- . (1981). Echelles et regles. In D. Sankoff & H. Cedergren (eds.), *Variation omnibus*. Carbondale, IL: Linguistic Research, Inc. 257–269.
- Sigley, Robert. (2003). The importance of interaction effects. *Language Variation and Change* 15:227–253.
- Swets, John A. (1998). Separating discrimination and decision in detection, recognition and matters of life and death. In D. Scarborough & S. Sternberg (eds.), *An invitation to cognitive science*. Vol. 4. *Methods, models and conceptual issues*, 2nd ed. Cambridge, MA: MIT Press. 635–702.
- Weinreich, Uriel, Labov, William, & Herzog, Martin. (1968). Empirical foundations for a theory of language change. In W. Lehman & Y. Malkiel (eds.), *Directions for historical linguistics*, 95–188. Austin: University of Texas Press. 95–188.