Soft Constraints Mirror Hard Constraints: Voice and Person in English and Lummi[1]

JOAN BRESNAN, SHIPRA DINGARE, AND CHRISTOPHER D. MANNING

Stanford University

Proceedings of the LFG 01 Conference

The University of Hong Kong, Hong Kong

Miriam Butt and Tracy Holloway King (Editors)

2001

CSLI Publications

http://csli-publications.stanford.edu/

[1]This work is based in part on research supported by the National Science Foundation under Grant No. BCS-9818077 and by a gift to the Stanford Symbolic Systems Program by Ric Weiland. This work omits some material in the LFG 01 conference presentation that will appear in a longer work in preparation by the authors.

ABSTRACT

The same categorical phenomena which are attributed to hard grammatical constraints in some languages continue to show up as statistical preferences in other languages, motivating a grammatical model that can account for soft constraints.

The effects of a hierarchy of person (1st, 2nd > 3rd) on grammar are categorical in some languages, most famously in languages with inverse systems, but also in languages with person restrictions on passivization. In Lummi, for example, the person of the subject argument cannot be lower than the person of a nonsubject argument. If this would happen in the active, passivization is obligatory; if it would happen in the passive, the active is obligatory (Jelinek and Demers 1983). These facts follow from the theory of harmonic alignment in OT: constraints favoring the harmonic association of prominent person (1st, 2nd) with prominent syntactic function (subject) are hypothesized to be present as subhierarchies of the grammars of all languages, but to vary in their effects across languages depending on their interactions with other constraints (Aissen 1999). There is a statistical reflection of these hierarchies in English. The same disharmonic person/argument associations which are avoided categorically in languages like Lummi by making passives either impossible or obligatory, are avoided in the SWITCHBOARD corpus of spoken English by either depressing or elevating the frequency of passives relative to actives. The English data can be grammatically analyzed within the stochastic OT framework (Boersma 1998, Boersma and Hayes 2001) in a way which provides a principled and unifying explanation for their relation to the crosslinguistic categorical person effects studied by Aissen (1999).

1 Categorical Effects of Person on Voice

The effects of the person hierarchy (1) on grammar are categorical in some languages, most famously in languages with inverse systems, but also in languages with person restrictions on passivization. In Lummi (Straits Salish, British Columbia), for example, the person of the subject argument cannot be lower than the person of a nonsubject argument. If this would happen in the active, passivization is obligatory; if it would happen in the passive, the active is obligatory (Jelinek and Demers 1983, 1994).[2]

(1) Person hierarchy:
1st, 2nd ≻ 3rd
(grammatical persons 'local' to speech act outrank others)

(2) Lummi examples:

a. *__ 'The man knows me/you'

b.

x̣či-t-ŋ=sən/=sx^{w}	ə	cə	swəyʔqəʔ
know-TR-PASS=1.SG.NOM/=2.SG.NOM	by	the	man

'I am/you are known by the man'

c.

x̣či-t-s	cə	swəyʔqəʔ	cə	swiʔqoʔəł
know-TR-3.TR.SUBJ	the	man	the	boy

'The man knows the boy'

d.

x̣či-t-ŋ	cə	swiʔqoʔəł	ə	cə	swəyʔqəʔ
know-TR-PASS	the	boy	by	the	man

'The boy is known by the man'

e.

x̣či-t=sən/=sx^{w}	cə	swəyʔqəʔ
know-TR=1.SG.NOM/=2.SG.NOM	the	man

'I/you know the man'

f. *__ 'The man is known by me/you'

The same holds in other, unrelated languages such as the Tanoan language Picurís (New Mexico) (Zaharlick 1982, Mithun 1999: 226–228) and the Southern Wakashan language

[2] The 'transitive' stem suffix *-t*, glossed TR, is one of a set that marks degree of volitionality of control of the action; the passive suffix -ŋ, glossed PASS, also marks middles (Jelinek and Demers 1994: 706). With local person arguments the active is obligatory. The Lummi pattern holds for bound pronouns; full nominal phrases designating speaker and hearer are formally 3rd person deictic expressions (Jelinek and Demers 1983: 173; 1994: 714).

Nootka (British Columbia) (Whistler 1985, Emanatian 1988). Although person-driven passives are sometimes viewed as inverses (cf. Klaiman 1991, Jacobs 1994, Forrest 1994, Jelinek and Demers 1983, 1994 on Salish), the syntactic contrasts in (3) have been drawn between person-driven passives and the Algonquian-type inverse exemplified by Plains Cree (Dahlstrom 1984), from Mithun (1999: 222–228):

(3)

Passive:	Inverse:
intransitive	transitive
patient Subject	patient Object
oblique case marking on agent	non-oblique agent
omissibility of indefinite agent	non-omissibility

On the basis of such contrasts, we accept with Mithun the evidence for the existence of person-driven passives. (Mithun 1999: 227 concludes of Picurís, "There is no question that these constructions are formally passive.")

2 A Theory of Passivization in Optimality Theory

From a classical generative point of view, it is difficult to see why person and voice should interact: after all, the person of arguments has nothing to do with verbal lexical semantic structure or case frames or the syntactic classification of lexical argument roles in LFG's lexical mapping theory (Bresnan 2001: ch. 14 and references), which have usually been taken to drive passivization. But from a different perspective, such interactions are not surprising.

It is well known that passives have properties of syntactically 'marked' constructions (Greenberg 1966, Trask 1979):

(i) Typological distribution: There are many languages without passives.

(ii) Language-internal distribution: Where it occurs, the passive is often more restricted than the active. For example, many languages restrict the passive agent (it may not appear, or may appear only in certain persons); others have a morphologically defective passive paradigm (lacking certain tenses, etc); only subclasses of active transitive verbs may passivize.

(iii) Morphological marking: Passivization is morphologically marked (Haspelmath 1990).

But why should this be? An historical explanation is that actives are basic (unmarked) verb types; passives arise from originally deverbal constructions such as stative adjectives or nominals by a historical process of verbalisation (Trask 1979, Estival and Myhill 1988, Haspelmath 1990, Garrett 1990). But the historical explanation does not answer the question: *Why* are actives the basic/unmarked verb types, rather than passives?

The intuition shared by many linguists and adopted by Aissen (1999) is that agents make better subjects than patients do. Semantically 'active' (proto-agent) arguments harmonically align with the most prominent syntactic argument positions; semantically 'inactive' (proto-patient) arguments harmonically align with the least prominent syntactic positions. The ultimate reasons for this alignment lie in the pragmatics of discourse and the cognitive biases of humans (see Givón 1979, 1983, Kuno and Kaburaki 1977, Kuno 1987, Ariel 1991, Warren and Gibson 2001, MacWhinney in progress for discussion of several proposals).

The detailed effects of harmonic alignment on grammars can be explicitly modelled in Optimality Theory (OT). In phonology, the sonority hierarchy aligns with syllable structure so that the most sonorous sounds are attracted to syllable peaks and the least sonorous sounds to syllable margins (see Kager 1999 for a synthetic overview). Aissen (1999) proposes that syntax is analogous, with the most agentive semantic arguments attracted to the structural 'peak' of the clause, the Subject, and the least agentive arguments to the non-Subjects. Harmonic alignment is formally defined for a binary scale and an n-ary scale (Prince and Smolensky 1993: 136). The binary scale refers to a structure (of the syllable or clause, for example), and the n-ary scale refers to a substantive dimension such as sonority or proto-agentivity. Harmonic alignment produces two aligned Harmonic scales, one showing how the elements of the n-ary scale are distributed with respect to the high element of the binary scale, the other showing how they are distributed with respect to the low element.

Aissen's (1999) theory of harmonic alignment in syntax is illustrated in (4).

(4)	Prominence scales:	Harmonically aligned scales:	OT constraint subhierarchies:
	S > O	$\mathrm{S}_{ag} \succ \mathrm{S}_{pt}$	$^*\mathrm{S}_{pt} \gg {}^*\mathrm{S}_{ag}$
	agent > patient	$\mathrm{O}_{pt} \succ \mathrm{O}_{ag}$	$^*\mathrm{O}_{ag} \gg {}^*\mathrm{O}_{pt}$

The prominence scales on the left are analogous to the structural hierarchy of the syllable and the sonority hierarchy, respectively.[3] The upper and lower ends of these prominence scales are harmonically aligned as shown in the middle. On the right these alignments are expressed in OT terms as subhierarchies of markedness constraints prefixed with the '*' ('avoid') operator and inverted so that the most disharmonic combinations will receive the worst constraint violations.

[3] Aissen (1999) assumes a binarized relational hierarchy, adopting the binary scale Subject > Nonsubject, which encapsulates both S > O and S > OBL. She also assumes a role hierarchy based on proto-role theory (Dowty 1991, Asudeh 2001).

Other constraints may be interleaved in constraint subhierarchies, enhancing or suppressing their effects, but the relative ranking of the constraints in a subhierarchy is fixed across languages.[4] Harmonic alignment of the person hierarchy (1) with the relational hierarchy (see n. 3) yields further constraint subhierarchies, which may interact with the harmonic alignments in (4):

(5) $*S_3 \gg *S_{1,2}$, $*O_{1,2} \gg *O_3$, $*Obl_{1,2} \gg *Obl_3$

The markedness of the passive compared to the active follows from the universal subhierarchy $*S_{pt} \gg *S_{ag}$. For a semantically transitive verbal input, the active and not the passive is the optimal expression, all else being equal:

(6)

	input: v(ag,pt)	$*S_{pt}$	$*S_{ag}$
	passive: S_{pt}, Obl_{ag}	*!	
☞	active: S_{ag}, O_{pt}		*

But if the active is the optimal expression of a semantically transitive input, how can passivization occur at all? The answer is of course that other constraints favor the passive: avoiding or 'backgrounding' the agent (Shibatani 1985, Thompson 1987), avoiding subjects newer than non-subjects in the discourse (Birner and Ward 1998), placing the topic in subject position to enhance topic continuity (Givón 1983, Thompson 1987, Beaver 2000), etc. Tableau (7) illustrates this outcome for English, taking Birner and Ward's theory that passive subjects tend to be discourse-newer than non-subjects as the basis for the constraint $*S_{newer}$, an instance of Aissen's (1999) schematic constraint $*S_x$.

(7) English avoids subjects newer than non-subjects ($*S_{newer}$):

	input: v(ag/new, pt)	$*S_{newer}$	$*S_{pt}$	$*S_{ag}$
	active: S_{ag},O_{pt}	*!		*
☞	passive: S_{pt},Obl_{ag}		*	

Tableau (8) illustrates this outcome for Lummi.[5]

(8) Lummi avoids third person subjects ($*S_3$):

	input: v(ag/3,pt/1)	$*S_3$	$*S_{pt}$	$*S_{ag}$
	active: S_{ag},O_{pt}	*!		*
☞	passive: S_{pt},Obl_{ag}		*	

[4]In syntactically ergative languages (Manning 1996), the preference for agentive subjects must be overridden.

[5]This analysis of Lummi differs somewhat from that given in Aissen (1999); it was derived from the Lummi data by the Gradual Learning Algorithm. See below for further discussion.

Crosslinguistic variation comes from reranking (see Aissen 1999 for details). In languages without passives, the constraint *S_{pt} is undominated by any of these countervailing constraints. In general, the same constraints are hypothesized to be present in all grammars, but are more or less active depending on their ranking relative to other constraints. Thus Lummi falls back on *S_{newer} with third person agent and patient:

(9)

	input: v(ag/3/new,pt/3)	*S_3	*S_{newer}	*S_{pt}	*S_{ag}
	active: S_{ag},O_{pt}	*	*!		*
☞	passive: S_{pt},Obl_{ag}	*		*	

And English suppresses the relation/person constraints (*S_3, etc.) by low ranking:

(10)

	input: v(ag/3, pt/1)	*S_{newer}	*S_{pt}	*S_{ag}	*S_3
☞	active: S_{ag},O_{pt}			*	*
	passive: S_{pt},Obl_{ag}		*!		

We know this because the disharmonic combinations are still grammatical in English, unlike Lummi: *She met me, She'll be met by you*.[6]

3 Statistical Person/Voice Interactions in English

In the OT framework of the present study, following Aissen (1999), the active and passive are viewed as alternative candidate expressions of the same input person/role combinations. Evidence of person/voice interactions in English has been given previously (Svartvik 1966, Estival and Myhill 1988; Seoane Posse 2000, DeLancey 1981, Kuno and Kaburaki 1977, Kuno 1987, cf. Kato 1979, reviewed in Dingare 2001). But for the present study what is needed is information about the systematic choices made rather than information about the distributions of subsets within passives or actives. Prior studies generally fail to provide the full joint distribution, from which we can reconstruct the conditional frequencies needed.[7] We have therefore examined the parsed SWITCHBOARD corpus, a carefully designed database of spontaneous telephone conversations spoken by over 500 American English speakers, both male and female, from a great variety of speech communities (Godfrey et al. 1992, Marcus et al. 1993). The conversations average 6 minutes in length, collectively amounting to 3 million words of text. We have used the parsed portion of this corpus, which contains 1 million words.

[6]We note that Aissen's (1999) constraint subhierarchy *$O_{ag} \gg$ *O_{pt}, taken as a whole, penalizes transitivity and would therefore favor passives over actives if ranked high enough. We defer discussion of this problem to future work.

[7]Estival and Myhill (1988) provide exactly the kind of information needed for animacy and definiteness, but they provide person frequencies only for the patient role.

We have found that the same disharmonic person/argument associations which are avoided categorically in languages like Lummi by making passives either impossible or obligatory, are avoided in the SWITCHBOARD corpus of spoken English by either depressing or elevating the frequency of passives relative to actives. Compared to the rate of passivization for inputs of third persons acting on third persons (1.2%), the rate of passivization for first or second person acting on third is substantially depressed (0%) while that for third acting on first or second (2.9%) is substantially elevated:

(11) English person/role by voice (full passives)

action:			# Act:	# Pass:	% Act:	% Pass:
1,2	$\rightarrow$	1,2	179	0	100.0	0.0
1,2	$\rightarrow$	3	6246	0	100.0	0.0
3	$\rightarrow$	3	3110	39	98.8	1.2
3	$\rightarrow$	1,2	472	14	97.1	2.9

The leftmost column in (11) gives the four types of inputs (local person acting on local, local acting on nonlocal, etc.). We estimate the number of times each input was evaluated as the number of actives plus passives with that person/structure association. We then calculate the rate of passivization as the number of times that input was realized as passive.

Though the percentage of full passives (with *by* phrases) in spoken conversational English is very small, the person/voice effects are highly significant (χ^2 = 115.8, $p < 0.001$; Fisher exact test, $p < 0.001$). Similar significance levels result if short passives are included, but we omit them because the person of the agent is not always clear. See Dingare (2001) for further analysis and detailed methodological discussion.

In sum, the 'hard' grammatical constraints on person/voice interactions seen in languages like Lummi, Picurís, and Nootka continue to show up as statistical preferences in English.

4 Why is English like Lummi?

It is "a mainstay of functional linguistics" that "linguistic elements and patterns that are frequent in discourse become conventionalized in grammar" (from a publisher's blurb on Bybee and Hopper 2001). On this view, Lummi is simply at an extreme point from English along the continuum of conventionalization that connects frequentistic preferences in usage to categorical grammatical constraints. But as noted by Bresnan and Aissen (to appear), it remains unclear in a conventional generative syntax by what mechanisms usage preferences can harden into grammatical conventions:

> Classical generative theories of formal grammar are designed with mathematically discrete and logically deterministic formal architectures. On these

theories, frequentistic processes (such as the conventionalization of usage preferences) must belong either to grammar-external 'performance' along with speech errors and memory limitations, or to external choices among competing dialect grammars. Yet neither of these alternatives is an adequate model of variation and change, as first pointed out by Weinreich, Labov, and Herzog (1968). The same is true of the variable effects of markedness hierarchies on syntax.

— Bresnan and Aissen (to appear)

Stochastic Optimality Theory offers a useful approach to this phenomenon. It is one of a family of new optimization-based theories of grammar that can provide a unified account of categorical, variable, and gradient data (see Anttila in press, Manning to appear, and references).

5 Stochastic Optimality Theory

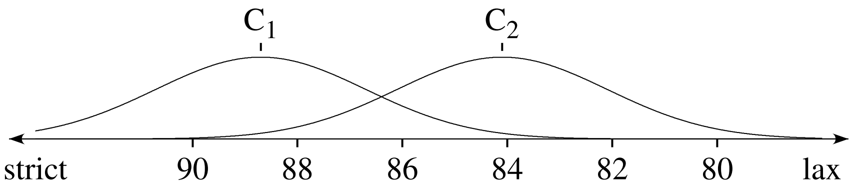
Stochastic OT (Boersma 1998, 2000, Boersma and Hayes 2001) differs from standard OT in two essential ways:

(i) **ranking on a continuous scale:** Constraints are not simply ranked on a discrete ordinal scale; rather, they have a value on the continuous scale of real numbers. Thus constraints not only dominate other constraints, but they are specific distances apart, and these distances are relevant to what the theory predicts.

(ii) **stochastic evaluation**: at each evaluation the real value of each constraint is perturbed by temporarily adding to its ranking value a random value drawn from a normal distribution. For example, a constraint with the mean rank of 99 could be evaluated at 98.12 or 100.3. It is the constraint ranking that results from these sampled values that is used in evaluation; it is referred to as the 'effective rank' here.

Figure 1 shows two constraints, C_1 and C_2. Note that the scale is inverted to match the standard OT convention that leftward is stronger in the constraint ranking. The ranks of these constraints are the means of their varying effective rankings, and are marked at the peaks of the bell curves; thus normally, $C_1 \gg C_2$. Nevertheless, on some evaluations the effective rank of C_1 will fall in the lower end of its normal distribution at the same time that the effective rank of C_2 falls in the higher end of its distribution, and a ranking reversal will occur, with $C_2 \gg C_1$. If C_1 and C_2 crucially conflict, such ranking reversals will create an alternative output for the same input, giving rise to variation.

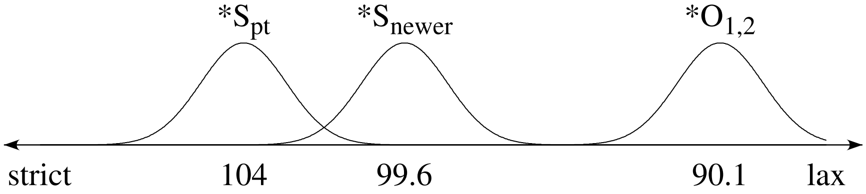
Figure 1: Constraint ranking on a continuous scale with stochastic evaluation



An OT grammar with stochastic evaluation can generate both categorical and variable outputs and can be learned from variable data by the GLA (Gradual Learning Algorithm, Boersma 1998, Boersma and Hayes 2001). Categorical outputs arise when crucially ranked constraints are distant. As the distance between constraints increases, interactions become vanishingly rare.[8] Variable outputs arise when crucially ranked constraints are closer together.

We can illustrate these ideas with the English 'pragmatic passive' grammar in Figure 2. The ranking values of the constraints (= the means of their normal distributions) are given on the X axis. The constraint *S_{pt} penalizes passives, but it is close enough to the constraint *S_{newer} to allow discernable variation. When ranking reversal occurs, as shown in the tableaux of (12), an alternative output occurs.

Figure 2: The English-type 'pragmatic passive'



[8] A distance of five standard deviations gives an expected reversal rate of less than 0.02% (Boersma and Hayes 2001: 50). Units of measurement are arbitrary. The standard deviation of ranking variation here is fixed at 2.0, so a ranking distance of 10 units between constraints is taken to be effectively categorical.

(12) Alternative outputs of the constraint ranking in Figure 2

	input: v(ag/new, pt)	$*S_{pt}$	$*S_{newer}$	$*O_{1,2}$
☞	active: S_{ag},O_{pt}		$*$	
	passive: S_{pt},Obl_{ag}	$*!$		

	input: v(ag/new, pt)	$*S_{newer}$	$*S_{pt}$	$*O_{1,2}$
	active: S_{ag},O_{pt}	$*!$		
☞	passive: S_{pt},Obl_{ag}		$*$	

In this stochastic grammar pragmatic or discourse-driven passivization is a statistical tendency, but not a categorical property of the output. Passives avoid subjects newer than non-subjects, but passivization is infrequent, and actives with new subjects also occur.

Where do the real number ranking values in a stochastic grammar come from? Starting from an initial state grammar in which all constraints have the same ranking values (arbitrarily set to be 100.0), the GLA is presented with learning data consisting of input-output pairs having the statistical distribution of, say, English. For each learning datum (a given input-output pair), the GLA compares the output of its own grammar for the same input; if its own output differs from the given output, it adjusts its grammar by moving all the constraints that disfavor its own output upward on the continuous ranking scale by a small increment, in order to make them apply more strictly, and moving all constraints that disfavor the given output downward along the scale by a small decrement, to relax their effects.[9] The adjustment process applies recursively to constraint subhierachies in order to preserve their local ordering relations.

6 Stochastic Grammars for English and Lummi

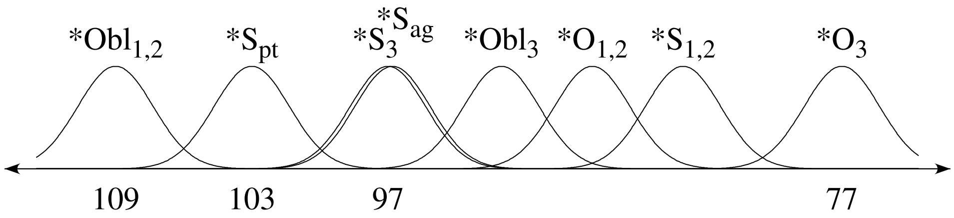
A partial stochastic grammar for the English person/voice interactions is given in Figure 3 and its output distribution in (13).

(13) Output distribution of the grammar in Figure 3

input:			% Active:	% Passive:
1,2	→	1,2	100.00	0.00
1,2	→	3	100.00	0.00
3	→	3	98.80	1.20
3	→	1,2	97.21	2.79

[9]The increment/decrement value is called the ‘plasticity’ and may be assumed to vary stochastically and to change with age (Boersma 2000).

Figure 3: Partial stochastic grammar of English



The constraint rankings and output distribution of this grammar were determined by simulation using the Praat system (Boersma and Weenink 2000), which includes an implementation of the Gradual Learning Algorithm.[10] Because of the stochastic components, the learned rankings and output distributions of grammars learned from the same distribution vary. The figures given are based on only one learned grammar; averaging over many grammars would better guarantee representativeness.

The constraints used are Aissen's (1999) constraint subhierarchies derived by harmonic alignment as outlined above. However, the grammar unrealistically omits the effects of the *S_{newer} constraint, which has a major influence on passivization in English (Birner and Ward 1998). Additionally, five of these constraints which are less active in our data were also omitted from the simulations for perspicuity: *Obl_{ag}, *Obl_{pt}, *O_{ag}, and *O_{pt}.

Observe that although the passive avoidance constraint *S_{pt} dominates the person-avoidance constraint *S_3, the two constraints are only 6 units apart (less than the near-categorical distance of 10; see n. 8), and will therefore produce low frequency variable outputs for some inputs. For inputs where only the agent is third person, passive outputs will occasionally be favored by *S_3, as shown in the tableau in (14):

(14) An (infrequent) effect of *S_3 on passive outputs:

	input: v(ag/3,pt/1)	*S_3	*S_{pt}	*S_{ag}
	active: S_{ag},O_{pt}	*!		*
☞	passive: S_{pt},Obl_{ag}		*	

[10]For learning the constraint rankings, a distribution of input-output pairs of person/voice combinations was specified according to the proportions given in our data in (11). The GLA learned directly from these distributions using the default settings in the Praat system for plasticity and replications. The relative ranking of constraints (the means) in the subhierarchies was maintained.

When both agent and patient are third person, the $*S_3$ constraint cannot decide between active and passive, and the decision passes to other constraints.[11]

The highest ranked constraint in Figure 3 is $*Obl_{1,2}$, which penalizes local person passive agents. It is more than 10 units (see n. 8) above any constraint that would disfavor an active (namely, $*O_{1,2}$ for an input with local-person patient and $*O_3$ for an input with third-person patient). These rankings reflect the zero frequency of local person passive agents in our data. Local person passive agents have been described as unacceptable in English (Kuno and Kaburaki 1977). Though they may be dispreferred or contextually marked, they are grammatical in spoken English. Among the examples cited by Kato (1979) are those in (15):

(15) a. I said, "Me watch it! Fuck that! Let him watch it." He was hired by me. I could fire him if I didn't like him. (Studs Terkel, *Working*)

b. When somebody says to me, "You're great, how come you're *just* a waitress?" *Just* a waitress. I'd say, "Why, don't you think you deserve to be served by me?" (Studs Terkel, *Working*)

With more training data and a more complete constraint set which includes factors of topicality and focus, our model should learn grammars that produce passives with local person agents. Note that if the ranking value of $*Obl_{1,2}$ in the grammar of Figure 3 were lowered from 109 to 104, the output of local person passives would increase to one-tenth of one percent, 0.1%, while barely changing the frequency of other outputs.

In sum, stochastic OT can capture the soft influence of person on English passivization that exists beneath the level of grammaticality judgments. Disharmonic person/argument combinations are grammatical but avoided, affecting the frequency of passivization.

Unfortunately we lack a parsed SWITCHBOARD corpus for Lummi. Nevertheless, it is possible to show by simulation how the descriptions of passive/voice interactions in Lummi grammar can also be captured by a stochastic OT grammar. We interpret the descriptions of Lummi from Jelinek and Demers (1983, 1994) by means of a simple distribution. Where a sentence type is described as ungrammatical, we assign it 0% relative frequency; where it is described as obligatory, we assign it 100%; and where it is described as optional, we assign it 50%:

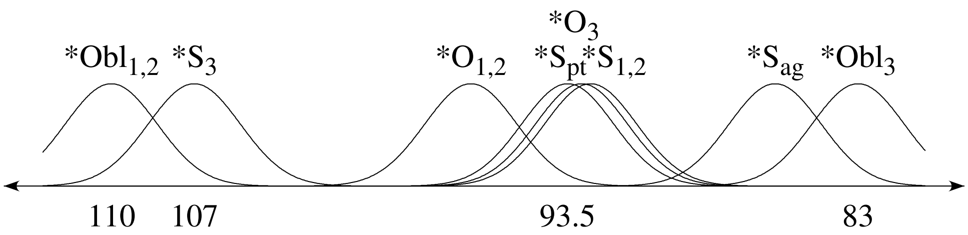
[11] For this input it will be $*S_{ag}$ that permits passive ouputs, with slightly less frequency than the passive outputs produced by $*S_3$, which is ranked marginally higher. In a less limited grammar other constraints would fill this role.

(16) Simulated Lummi (Straits Salish) input/output distribution

input:			% Active:	% Passive:
1,2	→	1,2	100.00	0.00
1,2	→	3	100.00	0.00
3	→	3	50.00	50.00
3	→	1,2	0.00	100.00

The simulated input/output distribution in (16) is then used to generate training data for the GLA, as before. The initial state of the grammar and the training regime (n. 10) are exactly the same as for English. A partial Lummi grammar learned by the GLA is shown in Figure 4.

Figure 4: Partial stochastic grammar of Lummi



Note that in contrast to the English grammar in Figure 3, the syntactic person-avoidance constraint *S_3 in the Lummi grammar is more than 10 units (5 standard deviations) above *S_{pt}. This ranking yields the obligatory passivization of inputs with local person patients and non-local person agents, capturing the categorical influence of person on Lummi passivization. The output distribution of the grammar in Figure 4 matches that in (16).

It might be thought that ranking on the continuous scale of real numbers is powerful enough to learn any distribution. In fact, under the present theory there are no stochastic OT grammars for 'anti-Lummi' or 'anti-English' distributions, which reverse the generalizations embodied in our data. Greater relative frequency of passives for first or second person acting on third would imply that third person subjects are avoided less than first or second person subjects. If so, then *$S_{1,2}$ must dominate *S_3 for a greater proportion of evaluations. But that ranking violates the subhierarchy in (5), which in stochastic OT requires the *mean* ranking values of these constraints to occur in the reverse order.

Thus, stochastic OT grammars are limited to subspaces of distributions that conform to the theory embodied in the constraint set. They are not general-purpose statistical analyzers and they have no special memory for frequencies (Boersma 2000).

7 Conventionalization and Frequency

Stochastic OT grammars allow us to place the person/voice interactions in English and Lummi at points on a continuum of conventionalization that connects frequentistic preferences in usage to categorical grammatical constraints. If this general perspective is correct, then we would expect to find languages at intermediate points on this same continuum. In the domain of person/voice interactions, Squamish (Coast Salish, British Columbia) may be a case in point.

Squamish and Lummi differ in their treatment of first person patient and nonlocal agent combinations: passivization is obligatory in Lummi but optional in Squamish; with second person patients passivization is obligatory for both languages. Within the ordinal OT framework, Aissen (1999) analyzes the difference between Lummi and Squamish by this constraint ranking:

(17) Lummi constraint ranking: $\ldots$ $*O_1, *O_2 \gg \ldots *S_{pt}$ — Squamish constraint ranking: $\ldots$ $*O_2 \gg \ldots *S_{pt} \gg *O_1 \ldots$

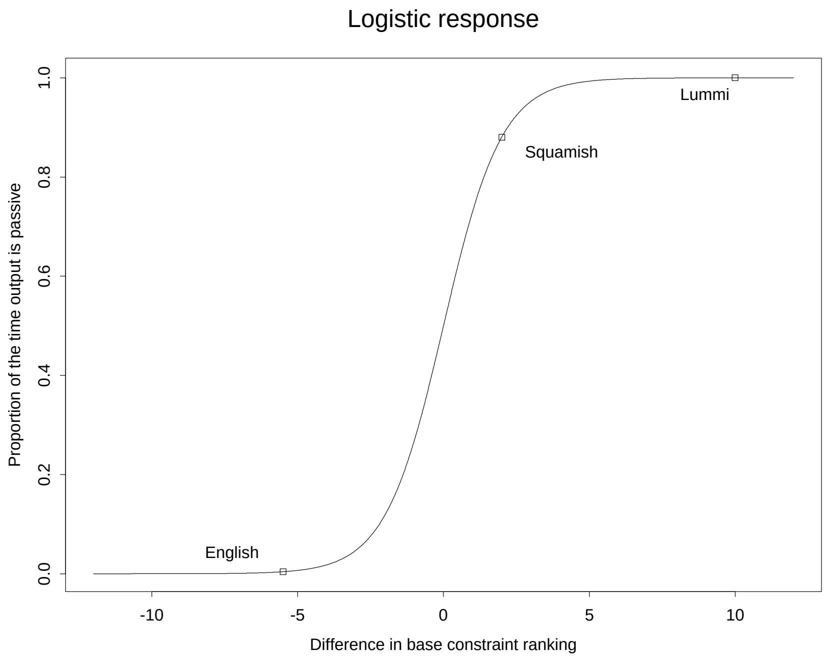
Given independent evidence that languages differ in whether first or second person is dominant (DeLancey 1981), Aissen assumes that the mutual ranking of the local-person avoidance constraints is not fixed by the subhierarchy, but subject to crosslinguistic variation.

However, it is not fully informative to say, as Aissen (1999) does, that passivization with third person agents and first person patients is optional in Squamish. In terms of what is preferred rather than what is merely possible, Squamish is described as being much the same as Lummi, "except that third person acting on first may be active, though commonly passive" (Klokeid 1969: 11). Thus in Squamish as in English, passives of the type *I was fooled by her* are optional alternatives to actives with disharmonic local-person objects: *She fooled me*. But in spoken English, as we have seen, such passives are exceedingly infrequent, far less common than the corresponding actives, while in Squamish they are more frequent than the corresponding actives.

In stochastic OT the rerankings postulated by Aissen (1999) take place on a continuous scale and imply *changes in frequency* as well as changes in grammaticality. The high rate of passivization with first person patients in Squamish shows that a constraint favoring passive, such as $*O_1$ is still ranked considerably above a constraint favoring active, such as $*S_{pt}$, on our continuous constraint ranking scale. Squamish and Lummi are closely related Coast Salish languages. It is plausible that we are observing a change in progress: the two languages represent different points in the changing categoricity of person effects on the passive, reflected in the ranking of the person-avoidance constraints for first and second person.

If a process of historical change is modeled by the movement in strength of a constraint along the continuous scale, as implied by the stochastic OT model, then (all else being equal) smooth changes in the relative frequencies of usage are predicted. Note however that although the change is smooth, it is not predicted to be linear. Rather, if a constraint reranking is crucial to the choice between two outputs, the prediction is that we should see an 'S' curve between the proportion of occurrences of the two outputs, of the sort that has been widely remarked on in historical and socio-linguistics (Weinreich, Labov, and Herzog 1968, Bailey 1973, Kroch 2001).

Figure 5: A conventionalization continuum.



More technically, assuming that the difference in ranking of two constraints which are crucial to the choice between two outputs A and B is changing linearly, then the proportion of output A is given by a logistic curve (Figure 5): when the constraints are at least 5 standard deviations apart, the proportion of the disfavored output is negligible; at 2 standard deviations, the rate reaches about 8%, but then it increases much more rapidly to 50% of each output when the two constraints are equally ranked.

These considerations suggest that classical grammatical descriptions in terms of what is 'possible' or 'grammatical' are overly idealized, concealing grammatically significant statistical structure beneath the idealization of linguistic intuitions of grammaticality.

8 Conclusion

The same categorical phenomena which are attributed to hard grammatical constraints in some languages continue to show up as statistical preferences in other languages, motivating a grammatical model that can account for soft constraints.

This observation is not new. Givón (1979: 26–31) already made this point forcefully over twenty years ago. What is new here is our demonstration that the stochastic OT framework can provide an explicit and unifying theoretical framework for these phenomena in syntax.

References

Aissen, Judith. 1999. Markedness and subject choice in Optimality Theory. *Natural Language & Linguistic Theory* 17: 673–711.

Anttila, Arto. In press. Variation and phonological theory. In J. Chambers, P. Trudgill and N. Schilling-Estes (eds.), *Handbook of language variation and change*. Oxford: Blackwell.

Asudeh, Ash. 2001. Linking, optionality, and ambiguity in Marathi. In P. Sells (ed.), *Formal and Empirical Issues in Optimality-Theoretic Syntax*, pp. 257–312. Stanford, CA: CSLI Publications.

Bailey, Charles-James N. 1973. *Variation and Linguistic Theory*. Arlington, Virginia: Center for Applied Linguistics.

Beaver, David. 2000. The optimization of discourse. Stanford: Stanford University Department of Linguistics MS. Draft on-line, Stanford University: http://www.stanford.edu/~dib/paper.htm.

Birner, Betty J. and Gregory Ward. 1998. *Information Status and Noncanonical Word Order in English*. Amsterdam: John Benjamins.

Boersma, Paul. 1998. *Functional Phonology. Formalizing the interaction between articulatory and perceptual drives*. Amsterdam: University of Amsterdam Ph.D. thesis.

Boersma, Paul. 2000. Learning a grammar in Functional Phonology. In J. Dekkers, F. v. d. Leeuw, and J. v. d. Weijer (eds.), *Optimality Theory: Phonology, Syntax, and Acquisition*, pp. 465–523. Oxford: Oxford University Press.

Boersma, Paul and Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32: 45–86.

Boersma, Paul and David Weenink. 2000. Praat computer program. On-line, Institute of Phonetic Sciences, University of Amsterdam: http://www.fon.hum.uva.nl/praat/.

Bresnan, Joan and Judith Aissen. To appear. Optimality and functionality: Objections and refutations. *Natural Language & Linguistic Theory* 20.1 Draft on-line, Stanford University: http://www-lfg.stanford.edu/bresnan/download.html.

Bresnan, Joan. 2001. *Lexical-functional Syntax*. Oxford: Blackwell.

Bybee, Joan and Paul Hopper (eds). 2001. *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins.

Dahlstrom, Amy. 1984. *Plains Cree Morphosyntax*. Berkeley: University of California, Berkeley Ph.D. dissertation. Published in 1991. New York: Garland.

DeLancey, Scott. 1981. An interpretation of split ergativity and related patterns. *Language* 57: 626–57.

Dingare, Shipra. 2001. *The effect of feature hierarchies on frequencies of passivization in English*. Stanford, California: Stanford University Department of Linguistics Master's thesis. On-line, Rutgers Optimality Archive: http://ruccs.rutgers.edu/roa.html. ROA-467-0901.

Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67: 547–619.

Emanatian, Michele. 1988. The Nootka passive revisited. In W. Shipley (ed.), *In Honor of Mary Haas. From the Haas Festival Conference on Native American Linguistics*, pp. 265–91. Berlin: Mouton de Gruyter.

Estival, Dominique and John Myhill. 1988. Formal and functional aspects of the development from passive to ergative systems. In Shibatani (ed.), pp. 441–91.

Forrest, Linda B. 1994. The de-transitive clauses in Bella Coola: Passive vs. inverse. In Givón (ed.), pp. 147–68.

Garrett, Andrew. 1990. The origin of NP split ergativity. *Language* 66: 261–296.

Givón, Talmy. 1979. *On Understanding Grammar*. New York: Academic Press.

Givón, Talmy (ed.) 1983. *Topic Continuity in Discourse. A Quantified, Text-based Study.* Amsterdam: John Benjamins.

Givón, Talmy. (ed.) 1994. *Voice and Inversion*. Amsterdam: Benjamins.

Godfrey J., E. Holliman and J. McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. *Proceedings of ICASSP-92*, San Francisco, pp. 517–520.

Greenberg, Joseph. 1966. *Language Universals. With Special Reference to Feature Hierarchies.* The Hague: Mouton.

Haspelmath, Martin. 1990. The grammaticalization of passive morphology. *Studies in Language* 14: 25–72.

Jacobs, Peter. 1994. The inverse in Squamish. In Givón (ed.), pp. 121–45.

Jelinek, Eloise and Richard Demers. 1983. The agent hierarchy and voice in some Coast Salish languages. *International Journal of American Linguistics* 49: 167–85.

Jelinek, Eloise and Richard Demers. 1994. Predicates and pronominal argments in Straits Salish. *Language* 70: 697–736.

Kato, Kazuo. 1979. Empathy and passive resistance. *Linguistic Inquiry* 10: 149–52.

Klaiman, M. H. 1991. *Grammatical Voice*. Cambridge: Cambridge University Press.

Klokeid, Terry J. 1969. Notes on the comparison of Wakashan and Salish. *University of Hawaii Working Papers in Linguistics* 7: 1–19.

Kroch, Anthony. 2001. Syntactic change. In M. Baltin and C. Collins (eds.), *The Handbook of contemporary syntactic theory*, pp. 699–729. Oxford: Blackwell Publishers.

Kuno, Susumu. 1987. *Functional Syntax*. Chicago: University of Chicago Press.

Kuno, Susumu and Etsuko Kaburaki. 1977. Empathy and syntax. *Linguistic Inquiry* 8: 627–72.

Manning, Christopher. 1996. *Ergativity: Argument Structure and Grammatical Relations*. Stanford, California: CSLI Publications.

Manning, Christopher. To appear. Probabilistic approaches to syntax. In R. Bod, J. Hay and S. Jannedy (eds.), *Probability Theory in Linguistics*. Cambridge, Massachusetts: The MIT Press. Draft on-line, Stanford University: http://www-nlp.stanford.edu/∼manning/papers/.

Marcus, Mitchell, Beatrice Santorini, and May Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn treebank. *Computational Linguistics* 19: 313–330.

Mithun, Marianne. 1999. *The Languages of Native North America*. Cambridge: Cambridge University Press.

Plank, Frans. (ed.) 1979. *Ergativity. Towards a Theory of Grammatical Relations*. New York: Academic Press.

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

Seoane Posse, Elena. 2000. The passive as an object foregrounding device in Early Modern English. In R. Bermudez-Otero, D. Denison, R. M. Hogg, C. B. McCully (eds.), *Generative Theory and Corpus Studies: A Dialogue from 10 ICEHL*, pp. 211–32. Berlin: Mouton de Gruyter.

Shibatani, Masayoshi. 1985. Passives and related constructions: A prototype analysis. *Language* 61: 821–48.

Shibatani, Masayoshi (ed.). 1988. *Passive and Voice*. Amsterdam: Benjamins.

Svartvik, Jan. 1966. *On Voice in the English Verb*. The Hague: Mouton.

Trask, R. L. 1979. On the origin of ergativity. In Plank (ed.), pp. 385–404.

Thompson, Sandra A. 1987. The passive in English: A discourse perspective. In R. Channon and L. Shockey (eds.), *In Honor of Ilse Lehiste*, pp. 497–511. Dordrecht: Foris Publications.

Warren, Tessa and Edward Gibson. 2001. The influence of referential processing on sentence complexity. Cambridge, Massachusetts: MIT Department of Brain and Cognitive Sciences MS. Submitted for publication.

Weinreich, Uriel, William Labov, and Marvin I. Herzog. 1968. Empirical foundations for a theory of language change. In W. P. Lehmann and Y. Malkiel (eds.), *Directions for historical linguistics. A symposium*, pp. 95–188. Austin: University of Texas Press.

Whistler, Kenneth W. 1985. Focus, perspective, and inverse person marking in Nootkan. In J. Nichols and A. C. Woodbury (eds.), *Grammar Inside and Outside the Clause. Some Approaches to Theory from the Field*, pp. 227–65. Cambridge: Cambridge University Press.

Zaharlick, Amy. 1982. Tanoan studies: Passive sentences in Picurís. *Ohio State University Working Papers in Linguistics* 26: 34–48.