The Stochastic Generalization
(Part III)

JOAN BRESNAN
Stanford University

[based on work by Bresnan, Dingare, and Manning]

[Optimality Theory and Typology, Summer School 2002]

1

I. The Generalization across Hard and Soft Constraints

Givón:

"What we are dealing with is apparently the very same *communicative tendency*—to reserve the subject position in the sentence for the *topic*, the old-information argument, the "continuity marker." In some languages (Krio, etc.), this communicative tendency is expressed at the *categorial* level of 100%. In other languages (English, etc.) the very same communicative tendency is expressed "only" at the *noncategorial* level of 90%. And a transformational–generative linguist will then be forced to count this fact as competence in Krio and performance in English."

— Givón (1979: 26–31)

2

Picurís examples:

Sənene mǫn-ʔąn.
man see-PAST
'The man saw him.' 3 → 3: passive optional

Mǫn-mia-ʔąn sənene-pa.
see-PASS-PAST man-OBL
'He was seen by the man.'

Sənene ti-mǫn-ʔąn.
man 1SG SUBJ.ANIM OBJ-see-PAST
'I saw the man.' 1,2 → 3: passive ungrammatical

**Sənene mǫn-mia-ʔąn nąʔpa.*
man see-PASS-PAST 1SG-OBL
'The man was seen by me.'

*__ 'The man saw me.' 3 → 1,2: passive obligatory

Ta-mǫn-mia-ʔąn sənene-pa.
$1\text{SG SUBJ}_{intrans}$-see-PASS-PAST man-OBL
'I was seen by the man.'

3

Lummi examples:

x̣čí-t-s cə swəyʔqəʔ cə swiʔqoʔəɬ
know-TR-3.TR.SUBJ the man the boy
'The man knows the boy' 3 → 3: passive optional

x̣čí-t-ŋ cə swiʔqoʔəɬ ə cə swəyʔqəʔ
know-TR-PASS the boy by the man
'The boy is known by the man'

x̣čí-t=sən/=sx^{w} cə swəyʔqəʔ
know-TR=1/2.SG.NOM the man
'I/you know the man' 1,2 → 3: passive ungrammatical

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

*__ 'The man knows me/you' 3 → 1,2: passive obligatory

x̣čí-t-ŋ=sən/=sx^{w} ə cə swəyʔqəʔ
know-TR-PASS=1/2.SG.NOM by the man
'I am/you are known by the man'

4

Evidence from English

(Bresnan, Dingare, and Manning's 2001 study of SWITCHBOARD)

Rate of Passivization

Agent ↓ Patient →	Local person	Third person
Local person	0.0%	0.0%
Third person	2.9%	1.2%

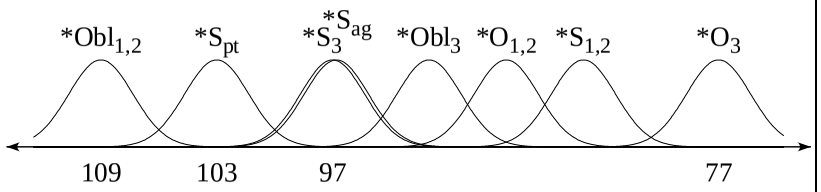
Compared to the rate of passivization for inputs of third persons acting on third persons (1.2%), the rate of passivization for first or second person acting on third is substantially depressed (0%) while that for third acting on first or second (2.9%) is substantially elevated.

Harmonic alignment gave us two particular hypotheses which are supported by these data: that the rate of passivization of 3 → 1,2 should be higher than for 3 → 3 (1-sided Fisher exact, $p < 0.008$); and that the rate of passivization of 1,2 → 3 should be lower than for 3 → 3 (1-sided Fisher exact, $p < 0.0001$).

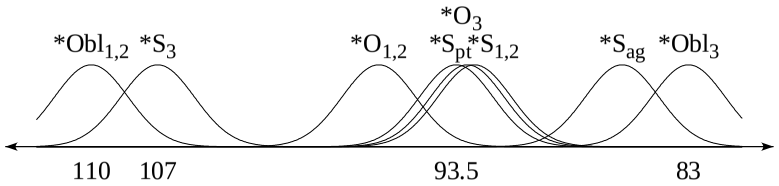
5

II. Stochastic OT Grammars

Partial stochastic grammar of English:

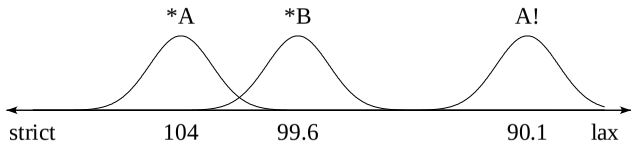


Partial stochastic grammar of Lummi:



6

The Gradual Learning Algorithm

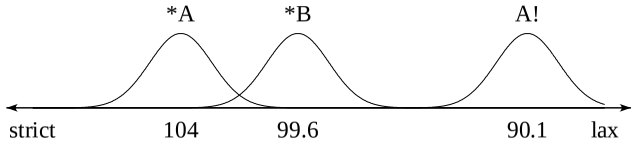


	*A	*B	A!
☞ cand1		*	
cand2	*!		

	*B	*A	A!
cand1	*!		
☞ cand2		*	

7

If cand_1 is correct, then when cand_2 is produced . . . :

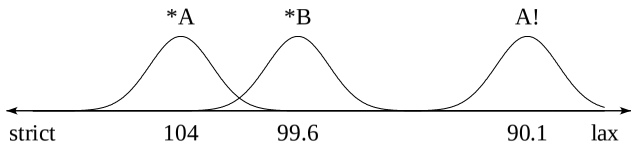


	*B⇒	⇐*A	A!
cand1	*!		
☞ cand2		*	

Result: *Categorical data repel constraints A* and B*.*

8

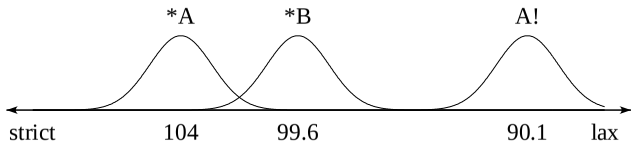
If cand_2 is correct, then when cand_1 is produced . . . :



	*A⇒	⇐*B	A!
☞ cand_1		*	
cand_2	*!		

Result: *Categorical data cause *A and B* to gradually rerank and then continue spreading apart.*

If both cand_1 and cand_2 are correct outputs for the same input, then . . . :



	*A⇒	⇐*B	A!
☞ cand_1		*	
cand_2	*!		

	*B⇒	⇐*A	A!
cand_1	*!		
☞ cand_2		*	

Result: *Variable data attract/repel constraints *A and *B into an eventual holding pattern that matches the frequency of variation.*

Why is English like Lummi and Picurís?

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 and Picurís are simply at an extreme point from English along the continuum of conventionalization that connects frequentistic preferences in usage to categorical grammatical constraints.

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.

Squamish example:

3 → 2: passive obligatory in Lummi and Squamish
3 → 1: passive obligatory in Lummi, optional in Squamish

Smooth Lummi-Squamish Reranking:

$$*\text{Obl}_{1,2} \gg *\text{S}_3 \gg *\text{O}_2, *\text{O}_1, *\text{S}_{pt}$$

Different points in the changing categoricity of person effects on the passive will be reflected by gradual changes in frequency, as the relative distance between constraints shrinks and grows.

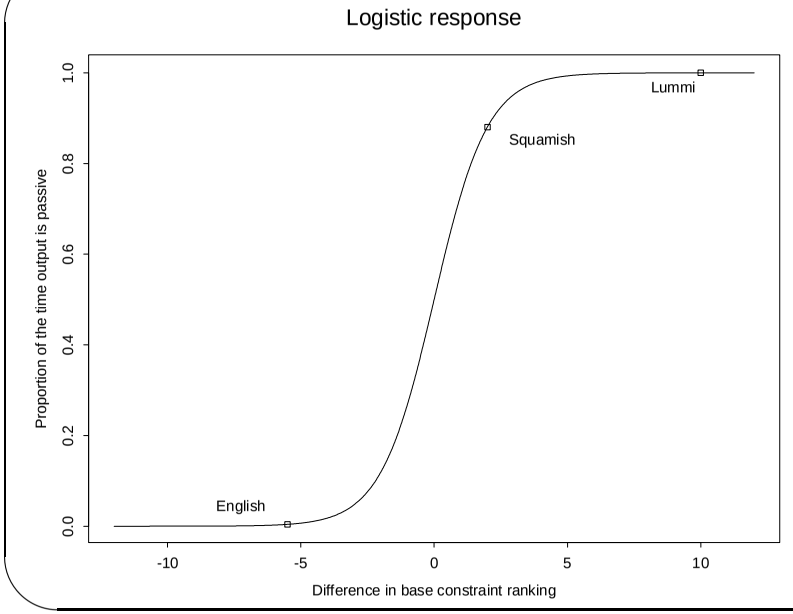
Reranking produces smooth changes in frequency—

If reranking is 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.

13

—but not linear changes:

If a constraint reranking is crucial to the choice between two outputs, and the distance between the two constraints is changing linearly, 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).



14

15

"Not all variability and heterogeneity in language structure involves change; but all change involves variability and heterogeneity."

— Weinreich, Labov, and Herzog (1968: 188)

16

III. Two theories of how the person hierarchy influences voice

perspective-based: empathy or perspective-taking (Kuno and Kaburaki 1977; DeLancey 1981; Kuno 1987; MacWhinney in progress, ao) — grammar is designed to facilitate perspective shifting during communication; interlocutors share the perspectives of speech-act participants and of referents having causal roles.

pragmatics-based: accessibility of referents in the pragmatic context (Givón 1976, 1979, 1994; Ariel 1991; Warren and Gibson 2001; cf. Gordon et al. 2001) — nominal expressions are most easily processed when their referents are contextually accessible

The connection to voice: Speech-act participants, referents having causal roles, and contextually accessible referents all tend to receive more attention and are consequently more frequently the subjects of predication.

Which is correct?

17

. . . perhaps both!

Constraints overlap:

local persons →
- humans
- expressed by pronominals
- given in the utterance situation (Speaker, Hearer)

third persons →
- nonhumans and humans
- expressed by nominals and pronominals
- not necessarily given in the utterance situation

Because of statistical dependencies of the input, other constraints can also derive a soft person effect.

18

Previous studies show that givenness/topicality interacts with the choice of active or passive. Among others:

Estival and Myhill (1988: 457–8):

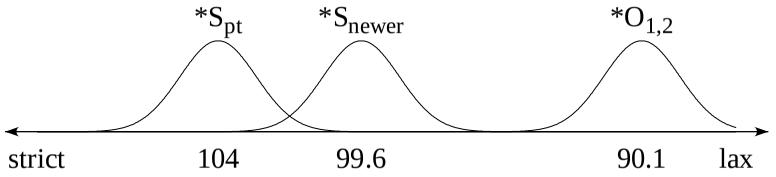
> The relative frequency of the choice of passive over active increases with 'nontopical' (nominal, nonhuman, and indefinite) agents and with 'topical' (pronominal, human, or definite) patients.

Weiner and Labov (1981: 46–54):

> The relative frequency of the choice of passive over active increases with given patients and decreases with new patients. ['Given' arguments are operationaly defined to have a coreferential NP within the preceding 5 clauses, irrespective of speaker turns.]

19

Hypothetical StOT grammar for the English-style 'pragmatic passive':[a]



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

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

[a] *S_{newer} = avoid subjects which are discourse-newer than non-subject arguments of the same clause (Birner and Ward 1998) (≈ Aissen's (1999) *Su/x).

20

How topicality/accessibility derives a soft person effect.

Given that first and second persons are seldom discourse new, while third persons may be (Cooreman 1987), one could assume that local person subjects are not penalized by the avoidance of newer subjects, while non-local person subjects are. Then a soft effect of the person hierarchy would follow from *S_{newer} in addition to any effect of person-avoidance constraints.

Nonlocal agents are differentially favored for passivization by newness (assuming that third person is inherently newer than local person):

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

input: v(ag/2, pt/3)	*S_{newer} (or *S_t)	*S_{pt}
☞ active: S_{ag},O_{pt}		
passive: S_{pt},Obl_{ag}	*!	*

21

Similarly, nonlocal patients are differentially disfavored for passivization by newness (again assuming that third person is inherently newer than local person):

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

	input: v(ag/3, pt/2)	*S_{newer} (or *S_t)	*S_{pt}
	active: S_{ag},O_{pt}	*!	
☞	passive: S_{pt},Obl_{ag}		*

22

Soft 'Topicality' Effects on English Active/Passive Choice: the Switchboard Corpus

The parsed Switchboard corpus is not tagged for topicality or givenness, but we can approximate this information-status concept by comparing the distributions of more and less definite nominal expression types, such as pronouns, proper names, definite and indefinite noun phrases.

In the Treebank Switchboard corpus local person pronouns are plentiful, but the distribution of pronouns and lexical (= nonpronominal) NPs is highly skewed (Francis et al. 1999):

- 91% of subjects are pronominal
- 66% of objects are lexical

We found significant 'topicality' effects on passivization in Switchboard, following the methods of Dingare (2001: 19–23).[a]

[a]The results reported here are work in progress by Bresnan, Dingare, and Manning, and should not be quoted without permission.

23

Switchboard Rates of Passive by Pronominality of Agent and Patient[a]

Pt → / Ag ↓	Pronoun	Non-Pronoun
Pronoun	A: 2457 P: 4 (0.16%)	A: 3527 P: 0 (0%)
Non-Pronoun	A: 179 P: 18 (9.14%)	A: 216 P: 8 (3.57%)

A: Active count; P: Passive count; (n%): percent passives

: one-sided Fisher exact, $p < 0.05$

: one-sided Fisher exact, $p < 0.0001$

[a]Pronouns = definite personal pronouns and reflexives; Non-Pronouns = Definites, Proper Names, and Indefinites.

24

Switchboard Rates of Passive by Definiteness of Agent and Patient[a]

Pt → / Ag ↓	All Definite	Indefinite
All Definite	A: 4756 P: 23 (0.48%)	A: 1567 P: 0 (0%)
Indefinite	A: 42 P: 7 (14.29%)	A: 14 P: 0 (0%)

A: Active count; P: Passive count; (n%): percent passives

: one-sided Fisher exact, $p < 0.05$

: one-sided Fisher exact, $p < 0.0001$

[a]All Definites = Pronouns, Proper Names, and Definites. Definites = NPs beginning with *the, this, that, these,* or *those* and not followed by a proper noun; Indefinites = NPs beginning with *a, an,* or *some* and not followed by a proper noun. NPs beginning with possessives are excluded.

Hard ‘Topicality’ Effects in Lummi

Lummi categorically avoids pronominal objects with non-pronominal subjects (Jelinek and Demers 1983, 1994):

*__ ‘The man knows it.’

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

‘It is known by the man’

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

‘He knows the man.’

Lummi categorically avoids indefinite subjects of transitive verbs (Jelinek and Demers 1994: 714, 732).

These and related facts[a] follow from the theory of harmonic alignment in OT: constraints favoring the harmonic association of referentially prominent arguments (pronoun, definite) with prominent syntactic function (subject) are hypothesized to be present as subhierarchies of the grammars of all languages, and to yield categorical grammaticality effects if they outrank other relevant constraints (Aissen 1999).

The stochastic generalization of OT explains how these categorical topicality effects of Lummi grammar can parallel the soft, frequentistic effects of ‘topicality’ on voice in English, which lie beneath the threshhold of grammaticality judgments.

[a]The avoidance of local person objects and passive agents does not extend to the free-standing referring expressions for local persons, which are focussed, hence ‘newer’ (Jelinek and Demers 1994: 714).

Recall Givón once again:

> “What we are dealing with is apparently the very same *communicative tendency*—to reserve the subject position in the sentence for the *topic*, the old-information argument, the “continuity marker.” In some languages (Krio, etc.), this communicative tendency is expressed at the *categorial* level of 100%. In other languages (English, etc.) the very same communicative tendency is expressed “only” at the *noncategorial* level of 90%. And a transformational–generative linguist will then be forced to count this fact as competence in Krio and performance in English.”
>
> — Givón (1979: 26–31)

Does that mean that person constraints are not needed to explain the interaction of person with voice?

No. Recall Squamish:

3 → 2: passive obligatory in Lummi and Squamish, optional in English
3 → 1: passive obligatory in Lummi, optional in Squamish and English

There is no independent reason to believe that the Speaker is systematically less ‘given’ than the Hearer in Squamish vs. Lummi, or that the Hearer is systematically more ‘given’ than the Speaker in Squamish vs. English.

Conclusion: person-avoidance is controlled independently of information structure.

Nevertheless, we expect them to overlap substantially . . . (Dingare 2001)

Statistical dependencies of the input cause overlapping constraints to rise and fall together under the Gradual Learning Algorithm.

The fact that two constraints have exactly the same violation marks on a given input means that during training, they will be treated the same (that is, demoted or promoted by the same amount). If the given input is frequent, a person constraint will end up close enough to an overlapping discourse constraint to drive the choice of the output to some degree even when the candidates are not driven by topicality. Hence, because a local person agent is rarely realized as an oblique, the speaker may disprefer passive even when the local person agent is non-topical.

Thus, because of the statistical dependencies of person and 'topicality' in the input, the person and discourse constraints will rise in tandem under the GLA. In the absence of active countervailing constraints, an emerging categoricity of person effects on voice will necessarily accompany an emerging categoricity of newness effects on voice (*S_{newer}), and vice versa.

29

IV. Questions about the Role of Frequency in Grammar

What does randomness really mean in a cognitive linguistic model?

The effective ranking ('selectionPoint') of a constraint C_i is given by the equation (Boersma 2000: 483):

$$selectionPoint_i = rankingValue_i + noise$$

The $noise$ variable represents unknown factors that are independent of the linguistic theory embodied in the constraint set. We assume that there is in fact a deterministic function from the total context plus the input to the output, but the context is too complex to know in detail. The random noise variable simply models our ignorance of the total context and of the non-linguistic factors that determine the probability of an output, for example by affecting the speaker's sensitivity to aspects of the current context.[a]

[a] To conclude that the variable usage modelled by this theory is therefore random and uncaused in the real world is the "fallacy of reified ignorance" described by Bresnan and Deo 2001.

30

Is all variation due to 'noise'?

No. Another determinant of frequency is style, which Boersma and Hayes (2001: 83–84) represent by a simple scale:

(maximally casual) $0 \leq Style \leq 1$ (maximally formal)

To model stylistic variation, they propose that the selection point for a given constraint C_i be determined by the following equation, where $styleSensitivity_i$ is a constraint-specific value:

$$selectionPoint_i = rankingValue_i + styleSensitivity * Style + noise$$

"Constraints with positive values for $styleSensitivity$ take on higher ranking vlaues in formal speech; constraints with negative values for $styleSensitivity$ take on higher ranking values in casual speech, and constraints with zero values of $styleSensitivity$ are style insensitive."

The style factor is not itself a grammatical constraint, but it boosts or depresses the rankings of groups of grammatical constraints in a systematic way which reflects a kind of sociolinguistic competence. Fluent, native knowledge of a language can be gauged precisely by the control of such factors and the ability to deploy them appropriately.

31

Does it make sense to derive frequencies of usage from grammar?

After all, unlike the grammaticality of a linguistic form, which is an idealization over usage, the actual frequency of usage of a form is a function of both grammatical structure (as determined by the theory embodied in the constraint set) and extra-grammatical factors such as memory limitations, processing load, and the context. These extra-grammatical factors are not represented by constraints in the stochastic grammars. Therefore the grammars that derive the given output distributions must be bogus, because their constraint rankings completely determine the distributions, when in fact non-constraint factors play an important role in determining frequency.

Knowledge of the grammatical structure of a particular language is represented by the (mean) ranking values of the constraints. Extra-grammatical factors affecting language use are represented by the variables that perturb the rankings. So each 'competence' grammar (= set of ranking values) is embedded in a 'usage' grammar (the style and noise variables). This embedding enables a much richer array of evidence to be used in studies of grammar than with classical approaches.

32

Doesn't this approach blur the line between competence and performance?

Data do not come into the world pre-theoretically classified as 'competence data' or 'performance data'. It is our *theories* which permit us to interpret some kinds of data and force us to disregard others.

As theory matures, the very same data are often reclassified. —Witness the development of modern semantic theory, which has brought more and more data earlier classified as 'pragmatic' and therefore outside of the bounds of grammar, within the scope of grammatical theory. Similarly, phonetics has increasingly come into the domain of recent phonological theory. Our study suggests that formal syntactic theory may be ripe for a similar development.

(Grammaticality judgments are just as much performances as more easily quantifiable behaviors. There are no privileged data for linguistic theory.)

Conclusion:

The same categorical phenomena which are attributed to hard grammatical competence constraints in some languages continue to show up as statistical preferences in other languages, motivating a grammatical model of competence that can account for soft constraints. We have shown how one example of this phenomenon can be successfully modeled in Stochastic Optimality Theory.

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.

References

Parts of this lecture are written up in the following, where references can be found:

Bresnan, Joan, Shipra Dingare and Christopher Manning. 2001. Soft constraints mirror hard constraints: Voice and person in English and Lummi. In M. Butt and T. H. King (eds.), *Proceedings of the LFG 01 Conference, University of Hong Kong*. On-line, CSLI Publications: http://csli-publications.stanford.edu/.

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

Assessment Question
Due Thursday, August 1st in class.

Prepare an essay of about 600 words on one of the two following questions.

Squib: Write a squib which brings some empirical data to bear on issues discussed in this course. Present the data carefully, and make clear how they are relevant to the themes of the class.

Read & Discuss: Read one of the following papers and focus on some argument or arguments that it develops. Summarize the argument clearly, and assess it, drawing on the class lectures and/or readings.

1. Martin Haspelmath (2001) "Explaining the ditransitive-role constraint: A usage-based approach." Can be downloaded from the class webpage.
2. Newmeyer, F. (2002). "Optimality and functionality: A critique of functionally-based optimality-theoretic syntax." *Natural Language & Linguistic Theory* 20(1): 43–80. Can be downloaded from: http://faculty.washington.edu/fjn/MH&FN_outline.html.

In connection with the Newmeyer paper, you may also want to consult:

Joan Bresnan and Judith Aissen (2002) "Optionality and functionality: Objections and refutations." *Natural Language & Linguistic Theory* 20(1): 81–95. Can be downloaded from the class webpage.