Typology in Variation:

A Stochastic Optimality-theoretic Approach to Person/Voice Interactions in English and Lummi (Straits Salish)

by

JOAN BRESNAN, SHIPRA DINGARE, AND CHRIS MANNING

Stanford University

Based on the slides for a Syntax Workshop given by Joan Bresnan

16 October 2001

In a nutshell . . .

i) The generalization: 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.

ii) A case study: The person hierarchy affects subject selection categorically in Lummi (Straits Salish, British Columbia), Nootka (Southern Wakashan, British Columbia), Picurís (Tanoan, New Mexico), and other languages. It also affects the frequency of subject selection in active/passive choices in spoken English.

iii) A model: We present a stochastic optimality theory account where the differences between Lummi and English are accounted for by positing different strengths for constraints within the same typologically motivated constraint system.

Cf. Givon on definiteness

"In many of the world's languages, probably in most, the subject of declarative clauses cannot be referential-indefinite .... In a relatively small number of the world's languages ... referential-indefinite nouns may appear as subjects of nonpresentative sentences .... When one investigates the text frequency of [such] sentences in English, however, one finds them at an extremely low frequency: About 10% of the subjects of main-declarative-affirmative-active sentences (nonpresentative) are indefinite, as against 90% definite. Now this is presumably not a fact about the "competence" of English speakers, but only about their actual "language behavior." But are we dealing with two different kinds of facts in English and Krio? Hardly. 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 tranformational-generative linguist will then be forced to count this fact as competence in Krio and performance in English. ... When live discourse data are taken into account ... it becomes obvious that noncategorial phenomena are the **rule** rather than the exception in human language." — Givón (1979: 26–31)

Outline:

a. The person hierarchy is rooted in psycholinguistic or communicative tendencies which affect not just the formal properties of a few particular languages, but every language.

b. The detailed effects of these tendencies on grammar can be captured in Optimality Theory (OT). The universal tendencies are modelled as violable constraints which have variable strengths (rankings) across languages. Given a language-particular ranking, an optimization function determines possible grammatical structures by minimizing the worst violations.

c. Frequentistic variation follows when these same constraints are ranked on a continuous scale with stochastic evaluation (Boersma 1998, 2000, Boersma and Hayes 2001). The resulting model defines a continuum of conventionalization which connects frequentistic preferences in usage to categorical grammatical rules.

The Person Hierarchy

1st, 2nd $\succ$ 3rd (local outranks nonlocal)

The Person Hierarchy

— appears at the top of a hierarchy of nominal features: e.g. 'animacy', 'topicality' hierarchies:

  1st, 2nd $\succ$ 3rd pronominal $\succ$ name $\succ$ human noun $\succ$ animate nonhuman noun $\succ$ inanimate noun

— ranks nominals according to their referents' "likelihood of participation in the speech event" (Smith-Stark 1974), their "inherent lexical content" (Silverstein 1976), their discourse-pragmatic topicality (Givón 1976, 1979, 1994), or their referents' accessibility during the psycholinguistic processing of language (Ariel 1990, Warren and Gibson 2001)

Why would the person hierarchy influence voice?

"Inherent lexical content" of nominals offers little insight.

Alternative theories:

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

pragmatics-based: accessibility of referents in the pragmatic context (Givón 1976, 1979, 1994; Ariel 1991; Warren and Gibson 2001) — nominal expressions are most easily processed when their referents are contextually accessible and their expressions occur in perceptually salient positions (e.g., **subject**) in linguistic structures

Categorical Effects of Person on Voice

The effects of the person hierarchy 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, 1994).

Lummi examples:

*__ ‘The man knows me/you’

x̣či-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’

x̣či-t-s cə swəyʔqəʔ cə swiʔqoʔəł
know-TR-3.TR.SUBJ the man the boy
‘The man knows the boy’

x̣či-t-N cə swiʔqoʔəł ə cə swəyʔqəʔ
know-TR-PASS the boy by the man
‘The boy is known by the man’

Lummi examples:

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

‘I/you know the man’

*__ ‘The man is known by me/you’

Like Lummi: Picurís (Tanoan, New Mexico) (Zaharlick 1982, Mithun 1999: 226–228) and Nootka (Southern Wakashan, British Columbia) (Klokeid 1969, Whistler 1985, Emanatian 1988).

Person-driven passives sometimes viewed as inverses (cf. Klaiman 1991, Jacobs 1994, Forrest 1994, Jelinek and Demers 1983, 1994 on Salish), but compare person-driven passives and the Algonquian-type inverse exemplified by Plains Cree (Dahlstrom 1984), from Mithun (1999: 222–228):

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

(Mithun 1999: 227 concludes of Picurís, “There is no question that these constructions are formally passive.”)

A Theory of Passivization in Optimality Theory

Passives are marked (Greenberg 1966, Trask 1979):

- Typological distribution: There are many languages without passives.
- Language-internal: 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.
- Morphological: Passivization is morphologically marked (Haspelmath 1990).

Why?

Historical explanation: actives are basic 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 why are actives the basic (unmarked) verb types, rather than passives?

The intuition: 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.

This harmonic alignment is ultimately to be explained by the underlying psycholinguistic and communicative theories.

It is represented in OT as an asymmetric partial ordering of constraints that share syntactic tier values.

The detailed effects of harmonic alignment on grammars can be explicitly modelled in 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's (1999) theory of harmonic alignment in syntax:

Prominence scales:

$\mathrm{S} \succ \mathrm{O}$

agent $\succ$ patient

Harmonically aligned scales:

$\mathrm{S}_{ag} > \mathrm{S}_{pt}$

$\mathrm{O}_{pt} > \mathrm{O}_{ag}$

OT constraint subhierarchies:

$*\mathrm{S}_{pt} \gg *\mathrm{S}_{ag}$

$*\mathrm{O}_{ag} \gg *\mathrm{O}_{pt}$

The relative ranking of the subhierarchy is fixed across languages, but interleaving of other constraints can alter its effects.[a]

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

Harmonic alignment of the person hierarchy with the relational hierarchy yields further constraint subhierarchies, which may interact with role-based harmonic alignment:

$$*S_3 \gg *S_{1,2}, \qquad *O_{1,2} \gg *O_3,$$

$$*Obl_{1,2} \gg *Obl_3$$

(This in turn follows from the perspective-based or the pragmatics-based theories of person/voice interaction.)

In OT, the universal subhierarchy *S_{pt} $\gg$ *S_{ag} entails the markedness of the passive compared to the active (all else being equal):

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

Then why is there passivization?

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.

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}		*	

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}		*	

In languages without passives, the constraint *S_{pt} is undominated by any of these countervailing constraints.

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.

Lummi falls back on *S_{newer} with third person agent and patient:

	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}	*		*	

English suppresses the syntactic person-avoidance constraints by low ranking:

	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*.

Statistical Person/Voice Interactions in English

To see if there is an effect of person on the selection of active or passive in English, we need information about the systemic choices made for each input. Prior studies generally fail to provide the full joint distribution, from which we can reconstruct the conditional frequencies needed.[a] We have therefore examined the parsed SWITCHBOARD corpus, a 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). The conversations average 6 minutes in length, collectively amounting to 3 million words. We have used the parsed portion of this corpus (Marcus et al. 1993), which contains 1 million words.

Although the frequency of passives is quite low in this corpus, the frequency of local pronouns is high.[b]

---

[a] 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.

[b] Francis et al. (1999) show that 91% of subjects in the parsed SWITCHBOARD corpus are pronominal.

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 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.

The person/voice effects are highly significant (χ^2_3 = 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.[a] Just the lower 4 cells give a significantly different rate of passive ($\chi^2 = 7.9$ ($p = 0.005$); Fisher's exact test $p = 0.012$).

[a] See Dingare (2001) for detailed analysis and methodological discussion.

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.

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.

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.

Why is English like Lummi?

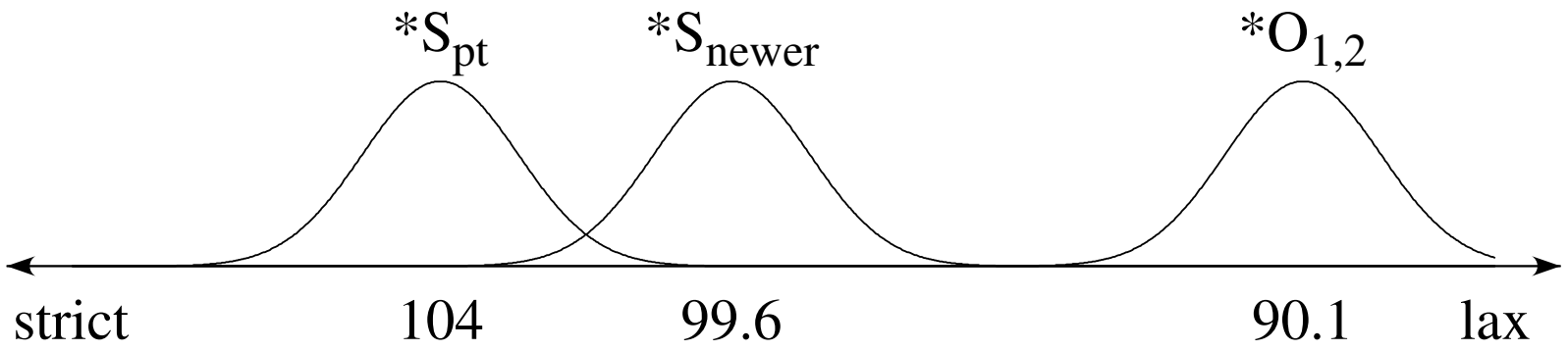
It is “a mainstay of functional linguistics” that “linguistic elements and patterns that are frequent in discourse become conventionalized in grammar” (from a publisher’s blurb on Bybee and Hopper 2001). On this view, Lummi is simply at an extreme point from English along the continuum of conventionalization that connects frequentistic preferences in usage to categorical grammatical constraints.

How can this process work in a conventional generative grammar?

There, 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 (Weinreich, Labov, and Herzog 1968).

Stochastic Optimality Theory

In Stochastic OT, variable outputs arise when crucially ranked constraints are close together. An example: the English-type 'pragmatic passive':

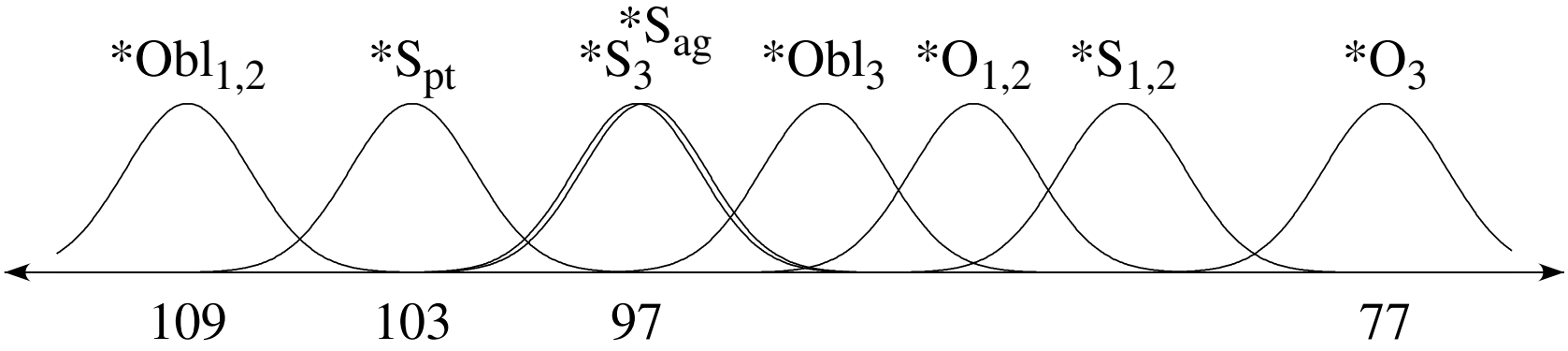


	input: v(ag/new, pt)	*S_{pt}	*S_{newer}	*$O_{1,2}$
☞	active: S_{ag},O_{pt}		*	
	passive: S_{pt},Obl_{ag}	*!		

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

Stochastic Grammars for English and Lummi

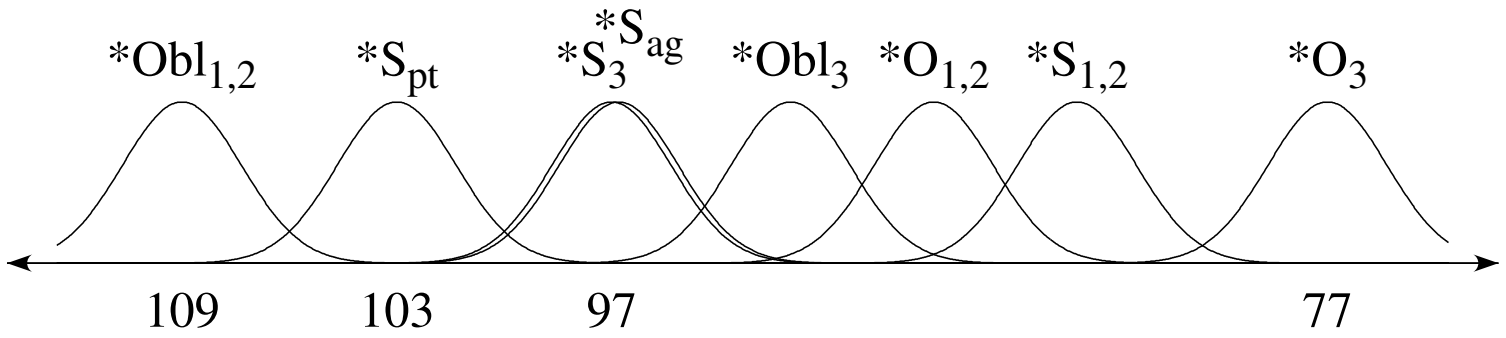
Partial stochastic grammar of English:



Output distribution of grammar:

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

again:



Observe: $*S_{pt} \gg *S_3$ but $|*S_{pt} - *S_3| = 6$, close enough to produce low frequency variable outputs for some inputs. For inputs where only the agent is third person, passive outputs will occasionally be favored by $*S_3$:

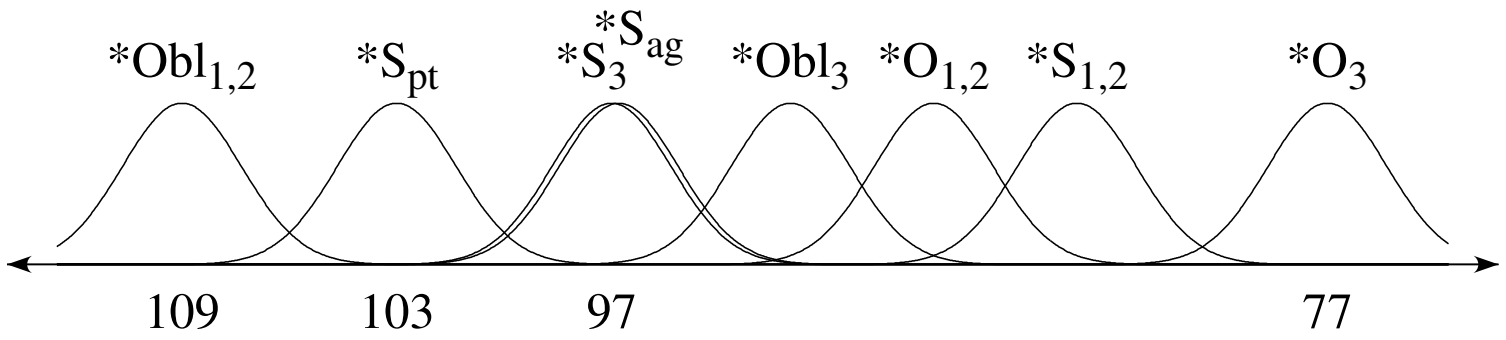
An (infrequent) effect of $*S_3$ on passive outputs:

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

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.

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. This grammar unrealistically omits the effects of the *S_{newer} constraint. Additionally, five constraints less active in our data were also omitted from the simulations for perspicuity: *Obl_{ag}, *Obl_{pt}, *O_{ag}, and *O_{pt}.

again:



Observe: $|$*Obl$_{1,2}$ $-$ *O$_{pers}|$ $>$ 10. (*O$_{1,2}$ disfavors an active 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. But Kato (1979) cites (from Studs Terkel, *Working*):

> 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.
>
> 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?"

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. (If the ranking value of *$\text{Obl}_{1,2}$ in the grammar 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.

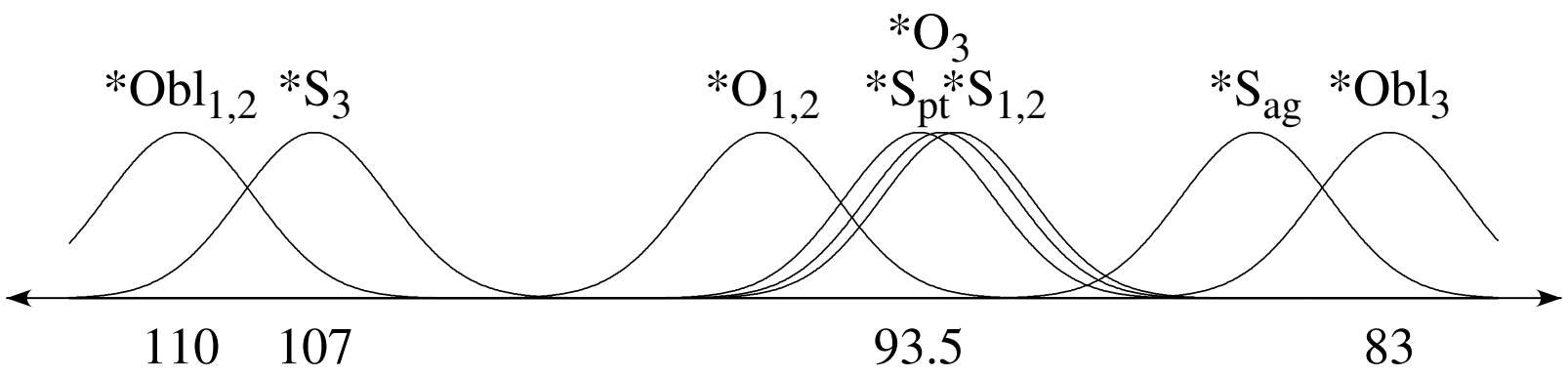
Unfortunately we lack a parsed SWITCHBOARD corpus for Lummi. Nevertheless, it is possible to show by simulation how the descriptions of passive/voice interactions in Lummi grammar can also be captured by a stochastic OT grammar. We interpret the descriptions of Lummi from Jelinek and Demers (1983, 1994) 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%:

Simulated Lummi input/output distribution:

input:			% Active:	% Passive:
1,2	$\rightarrow$	1,2	100.00	0.00
1,2	$\rightarrow$	3	100.00	0.00
3	$\rightarrow$	3	50.00	50.00
3	$\rightarrow$	1,2	0.00	100.00

The simulated input/output distribution is then used to generate training data for the GLA, as before. The initial state of the grammar and the training regime are exactly the same as for English.

Partial stochastic grammar of Lummi



Observe: $|*S_3 - *S_{pt}| > 10$. 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 matches the similated learning distribution exactly.

Isn't ranking on the continuous scale of real numbers powerful enough to learn any distribution?

No, it isn't. 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 constraint subhierarchy, which in stochastic OT requires the *mean* ranking values of these constraints to occur in the reverse order.

Thus, the output of stochastic OT grammars are limited to subspaces of distributions that conform to the theory embodied in the constraint set – the sharing of the effect of constraint violations across inputs, and in particular, here, the constraint subhierarchies. Within that feasible space, they can match input frequencies. But they are not completely general-purpose statistical analyzers and they do not just memorize frequencies (Boersma 2000).

Conclusion

We have argued that 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 phenomena can be successfully modeled in Stocastic Optimality Theory. Much work remains to apply these ideas to other phenomena.

References

Parts of this talk 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.