Questions . . .

[Optimality Theory and Typology, Summer School 2002]

1

I. The Generalization across Hard and Soft Constraints

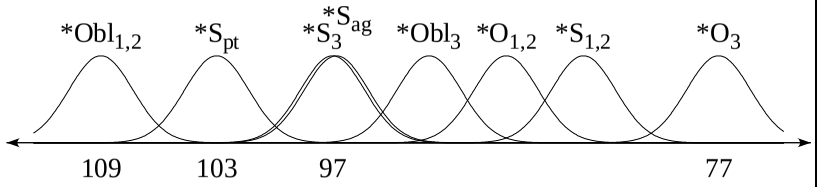
Givón:

> "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)

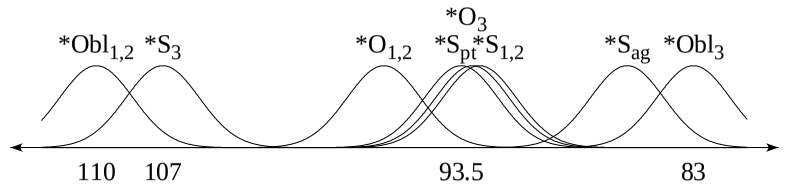
2

II. Stochastic OT Grammars

Partial stochastic grammar of English:



Partial stochastic grammar of Lummi:



3

Could competing conventional generative grammars explain the passive variation in English?

The competing grammars theory of variation is a model of diglossia (Kroch 2001). On the diglossic model of variation, the contact between two different populations having different grammars leads to internalization of competing grammars by individual speakers, who control two separate varieties. For example, some historical changes in English word order are attributed to the influence of Scandinavian speakers in Northern England (Kroch and Taylor 1997).

Could the diglossic model explain our passive findings? On this account, individual speakers would vary in the frequency of passive outputs because they have internalized alternative grammars which they deploy with varying frequency. The different grammars would have arisen from contact between different populations speaking varieties of English with and without the passive construction for certain person/role combinations. One population would have Lummi-like gaps in actives and passives as a hard constraint of their English grammar, perhaps as a result of some parameter setting of UG.

4

Some early studies propose that middle-class English speakers use an 'elaborated code' which has a higher proportion of passive verbs among all finite verbs than a 'restricted code' of working-class speakers, which has a lower percentage (Bernstein 1971 ao). But these studies have been criticized for failing to isolate the syntactic choice between active and passive, which shows no signficant difference between these groups (Weiner and Labov 1981: 32). (Passives should be compared to equivalent actives, rather than to all sentences. The latter can be influenced by differences in what is talked about, given that passives require fewer arguments than actives.)

5 Spontaneous speech shows significant stylistic and discourse effects on the choice of (agentless) passive or generalized-subject active.[a] But: "All of these conditions on the selection of active vs. passive are general features of the English language, used in much the same way by the very different sub-sections of the speech communities that we studied."(Weiner and Labov 1981: 56).

Conclusion: Diglossia is an unlikely model for our passive data. *"All sections of the population appear to treat the passive/active choice in the same way, and conversely, the same constraints are found throughout the speech community."* (Weiner and Labov 1981: 56)

[a]Generalized pronoun subjects ("they") are characteristic of colloquial English, while passives are a mark of formal scientific and literary discourse; passives are also favored by the discourse tendencies to preserve subject reference and structural parallelism.

Why should we believe that the same constraints are present in every grammar?

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

Is it necessary to assume the constraints are innate?

No, universality does not imply innateness.

6 Some constraints may reflect innate biases, some may reflect common circumstances of the pragmatic environment.

How can a speaker have a (non-innate) constraint, grounded or not, for which s/he has no evidence? (Newmeyer)

Constraints which are not domain-specific are evidenced independently of their role in grammar (which may be overridden by other constraints). For example, the shape of the human vocal tract makes certain sounds more difficult to produce irrespective of any distributional evidence for their presence or absence that a speaker may encounter in learning a particular language.

Two (broad) theories of the person hierarchy:

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.

7 **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.***

What is gained by the model?

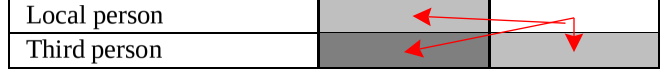
Recall: Logical Entailment of Implicational Universals

The theory of harmonic alignment logically entails certain crosslinguistic generalizations, which follow from the constraint subhierarchies and the transitivity of constraint domination (≫) in ordinal ('vanilla') OT.

Comrie (1989: 128): *"…the most natural kind of transitive construction is one where the A is high in animacy and definiteness and the P is lower in animacy and definiteness; and any deviation from this pattern leads to a more marked construction."*

8 *The spread of markedness:*

Agent ↓ Patient →	Local person	Third person
Local person		
Third person		



Disregarding other constraints, if passivization is categorical for some input, then it must be categorical for any more marked input (Dingare 2001: 16–17). For example, in Lummi and Picurís, passive is obligatory for input from the lower left cell and optional for input from the lower right cell. *Prediction: In no languages does the reverse hold.*

Generalization: Predictions of Relative Frequency

Disregarding other constraints, if passivization occurs with some *frequency* for a given input, then (by Aissen's theory of harmonic alignment expressed within the Stochastic OT model) it must occur with equal or higher *frequency* for any more marked input (Dingare 2001: 18).

Agent ↓ Patient →	Local person	Third person
Local person		
Third person		

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.

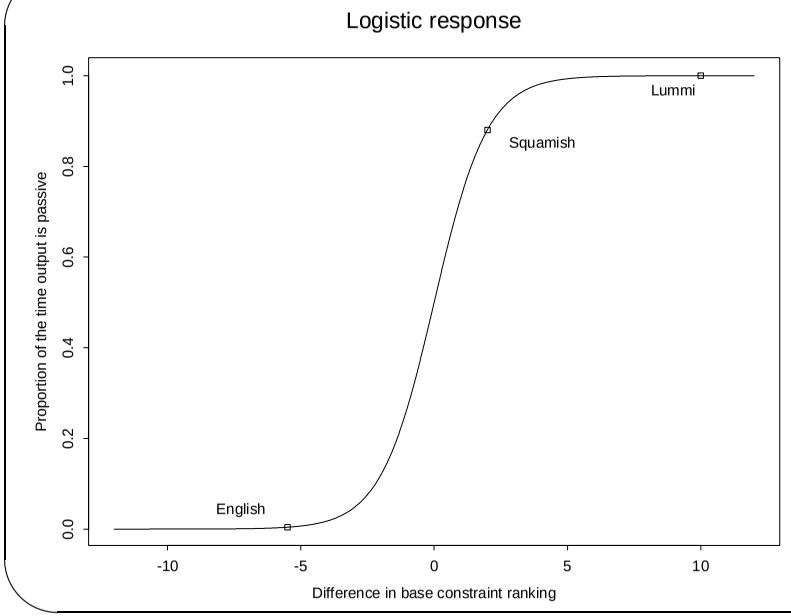
Reranking produces smooth changes in frequency—

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.

—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).

13



14

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

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

15

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.

16

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.

17

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.

18

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.)

19

Where is the rest of the grammar?

Answer. There are further lexical, morphosyntactic, and syntactico-semantic optimizations, for which we must choose a specific representational basis.

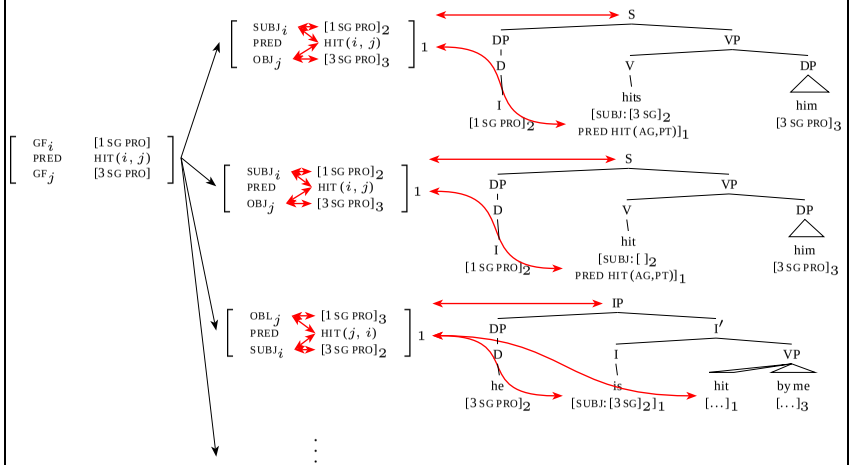
In the OT-LFG formalization (using LFG as the representational basis for OT syntax), these can be computed in parallel. Examples: verbal agreement choices, the selection of analytic or synthetic forms, and sentential word order patterns.[a]

Language-particular effects follow from anchoring general families of constraints to specific word classes, paradigms, or morphs.

[a] See Choi 1999, Bresnan 2000, 2001a,b,c, in press; Bresnan and Deo 2001; Kuhn 2001; Lee 2001, Clark 2002a,b, Sells 2001a,b, Koontz-Garboden 2002, and references.

20

Parallel Optimizations (OT-LFG):



How can language particularity be accounted for in OT?

- Systematic language-particular properties are derived by constraint ranking.
- Some universal constraint families are indexed to language-particular word classes or morphemes, such as the family of morphological alignment constraints (McCarthy and Prince 1993) and morphologically indexed faithfulness constraints (Urbanczyk 1995, 1996; Benua 1995, 1996; Fukazawa 1997).
- Unsystematic properties of a particular language must be lexically specified.