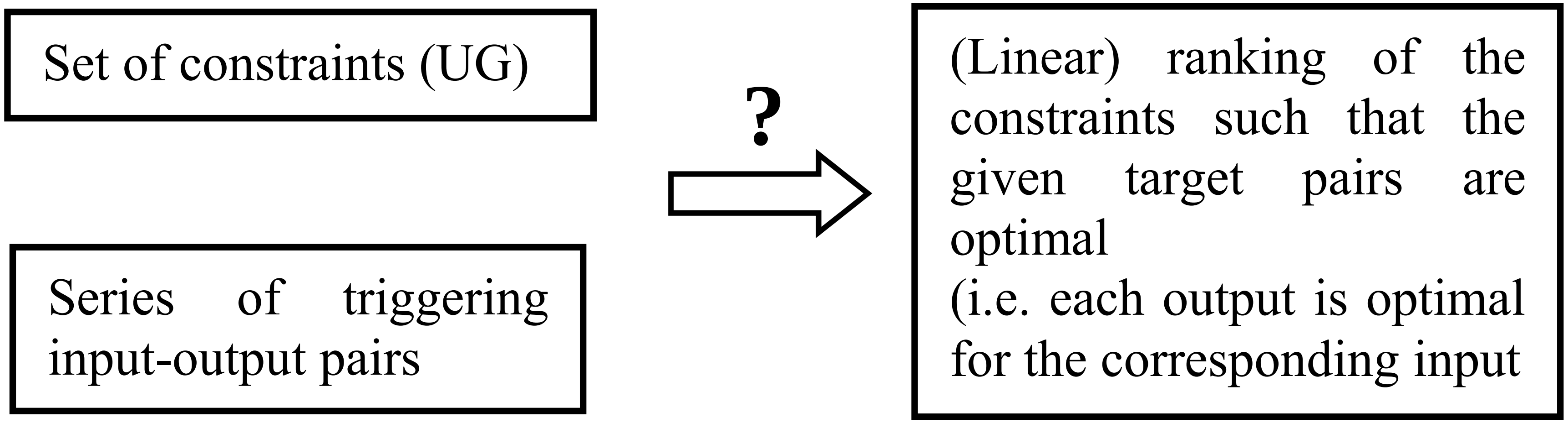
Lecture 3: OT learning Theory

1. Extracting constraint rankings from given input-output pairs
2. Constraint demotion: Prerequisites and formal results
3. The comprehension/production dilemma in child language
4. The OT learning algorithm
5. Richness of the base and constraints on inventories

1 Extracting constraint rankings from given input-output pairs



This is a simplified basic picture only: It is …

- useful for manually constructing grammars from given pairs
- requiring a list of relevant (hidden) inputs
- unrealistic as a model of language learning (inputs are hidden units – we have no direct access to them!)

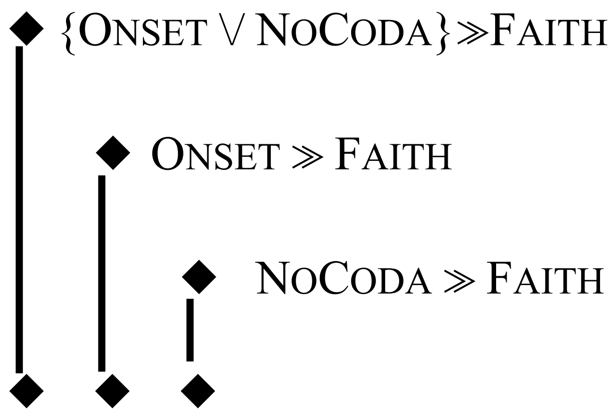
Example

Consider the basic syllable theory of the previous lecture with the system of constraints: {FAITH , ONSET, NOCODA}.
Extract the right ranking from input-output pairs like:

/atat/ - .a.tat.	(*English*)
/atat/ - .a.ta.⟨t⟩	(*Hawaiian*)
/atat/ - .□a.tat.	(*Yawelmani*)
/atat/ - .□a.ta.⟨t⟩	(*Senufo*)

optimal *Senufo*

Input: /atat/		FAITH	ONSET	NOCODA
1 (*English*)	.a.tat.		*	*
2 (*Hawaiian*)	.a.ta.⟨t⟩	*	*	
3 (*Yawelmani*)	.□a.tat.	*		*
✔ 4 (*Senufo*)	.□a.ta.⟨t⟩	**		



A candidate w is considered to be optimal iff for each competitor w', the constraints that are lost by w must be ranked lower than at least one constraint lost by w'.

Learning is assumed to be triggered by (positive) input-output pairs (which should come out as *grammatical* with regard to the "learned" Grammar). Each pair brings with it a body of negative evidence in the form of competitors (provided by **Gen**). This fact has to be emphasized as one of the main advantages of a connectionist theory like OT.

Input: /atat/		ONSET	NOCODA	FAITH
1 (*English*)	.a.tat.	*	*	
2 (*Hawaiian*)	.a.ta.⟨t⟩	*		*
3 (*Yawelmani*)	.□a.tat.		*	*
✔ 4 ☞ (*Senufo*)	.□a.ta.⟨t⟩			**

Information about the ranking, collected from 4≻3, 4≻2, 4≻1:

{ONSET, NOCODA} ≫ FAITH

Constraint demotion

Given a certain input *I* and a target output *SD*. The input is paired with a competitor *SD'*. This constitutes a Winner-Loser Pair: *SD* ≻ *SD'*.

For any constraint C which is lost by the winner SD, if C is not dominated by a constraint C' lost by the competitor SD', demote C to immediately below the highest constraint that is lost by SD'.

Example 1

Input: /atat/		FAITH	ONSET	NOCODA
1 (*English*)	.a.tat.		*	*
2 (*Hawaiian*)	.a.ta.⟨t⟩	*	*	
3 (*Yawelmani*)	.□a.tat.	*		*
✔ 4 (*Senufo*)	.□a.ta.⟨t⟩	**		

A sample run for *Senufo*:	start	{ONSET, NOCODA, FAITH}
	$4 \succ 3$:	{ONSET, NOCODA} $\gg$ FAITH
	$4 \succ 2$:	"
	$4 \succ 1$:	"

Example 2

Input: /atat/		FAITH	ONSET	NOCODA
1	.a.tat.		*	*
✔ 2	.a.ta.⟨t⟩	*	*	
3	.□a.tat.	*		*
4	.□a.ta.⟨t⟩	**		

A sample run for *Hawaiian*:

start: {ONSET, NOCODA, FAITH}
2 ≻1: {ONSET, NOCODA} ≫ FAITH
2 ≻3: NOCODA ≫ {FAITH, ONSET}
2 ≻4: NOCODA ≫ FAITH ≫ ONSET

2 Constraint demotion: Prerequisites and formal results

(A) UG = *Gen* + *Con.* The learning problem consists in inferring the ranking of the constraints in *Gen.* This excludes both the possibility that the constraints themselves are learned (in part at least) or that aspects of the generator are learnable.

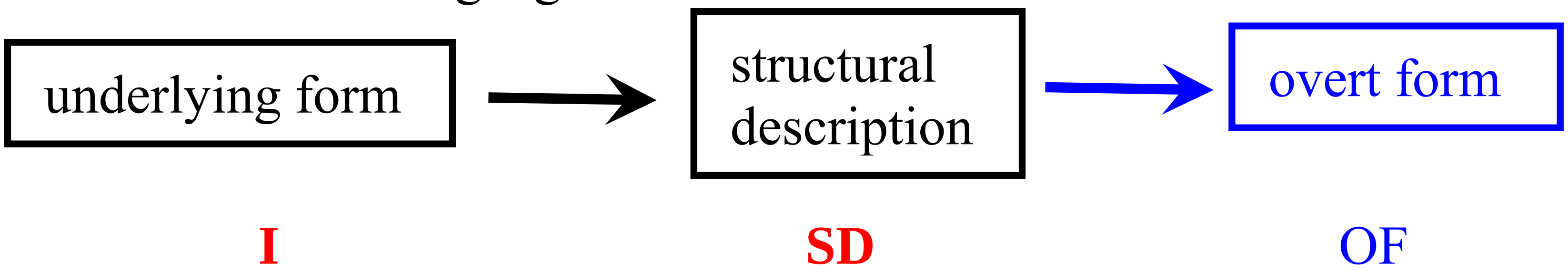
(B) The force of strict domination $\gg$: A relation of the form C $\gg$ C' does not merely mean that the cost of violating C is higher than that of violating C'; rather, it means that no number of C' violations is worth a single C violation. The force of strict domination excludes cumulative effects where many violations of lower ranked constraints may overpower higher ranked constraints.

(C) The OT grammar of the language that has to be learned is based on a *total* ranking of all the constraints: $C_1 \gg C_2 \gg \ldots \gg C_n$.

During learning the ranking of the constraints is not restricted to a total ranking. Instead, more general domination hierarchies are admitted which have the following general form:

$\{C_1,C_2, \ldots,C_3\} \gg \{C_4,C_5, \ldots,C_6\} \gg \ldots \gg \{C_7,C_8, \ldots,C_9\}$. ("stratified domination hierarchy")

(D) In the (theoretically) simplest case, learning is triggered by pairs <*I*, *SD*> consisting of a (hidden) input and a structural description *SD* of the source language *L*.



Fact 1: Correctness of iterative constraint demotion

The iterative procedure of constraint demotion converges to a set of totally ranked constraint hierarchies, each of them accounting for the learning data. Interestingly, this result holds when starting with an arbitrary constraint hierarchy. (cf. Tesar & Smolensky 2000)

Fact 2: Data complexity of constraint demotion

Consider a system with a fixed number of constraints, say N. The number of informative data pairs required for learning is no more than N(N-1)/2, independent on the initial hierarchy and the nature of the constraints. (cf. Tesar & Smolensky 2000)

Hint for proof: Crucial is the inherently comparative character of OT. Assuming N constraints, then for each pair $1 \leq i, j \leq N$ it has to be decided whether $C_i \gg C_j$ or $C_i \gg C_j$. There are exactly N(N-1)/2 such decisions and each one can be brought about on the basis of one appropriate data pair triggering the corresponding set of winner-loser pairs. Consequently, no more than N(N-1)/2 appropriate data pairs should be necessary for learning the correct grammar.

Comparison between OT and P&P

Let's assume a parameterized UG with n parameters. Then this system admits 2^n grammars when the parameters are binary. In the worst case, the average number of triggers before reaching the target grammar is 2^n. This is due to the fact that the learner is informed about the correct value of the different parameters by positive data only, and that all parameters are interacting in the worst case.

		Number of Grammars	Number of Triggers
P&P	30 binary parameters	$2^{30} = 1{,}073 \times 10^9$	$1{,}073 \times 10^9$
OT	20 constraints	$20! = 2{,}43 \times 10^{18}$	190

3 The comprehension/production dilemma in child language

Children's linguistic ability in production lags dramatically behind their ability in comprehension.

Standard reaction of Generative Grammar: dramatically greater competence-performance gap for children. Typically: children do not produce a particular segment because their motor control hasn't yet mastered. However, Menn & Mattei (1992) show that children who systematically avoid a given structure in their linguistic production can often easily imitate it.

Jacobson's generalization

The same configurations which are marked in the sense of disfavoured in adult languages tend also to be avoided in child language.

Consequence: constraints defining linguistic markedness are shared across adult and child language production. It would be attractive to have a viable hypothesis according to which **Grammar** has a central role to play in explaining child production.

The two horns of the dilemma

(1) Competence-performance gap for children (empirically wrong)

(2) Two grammars for children, one for production, the other for comprehension (extremely unattractive)

OT provides a simple way out of this dilemma. The point is that the structures that compete are different in production and comprehension!

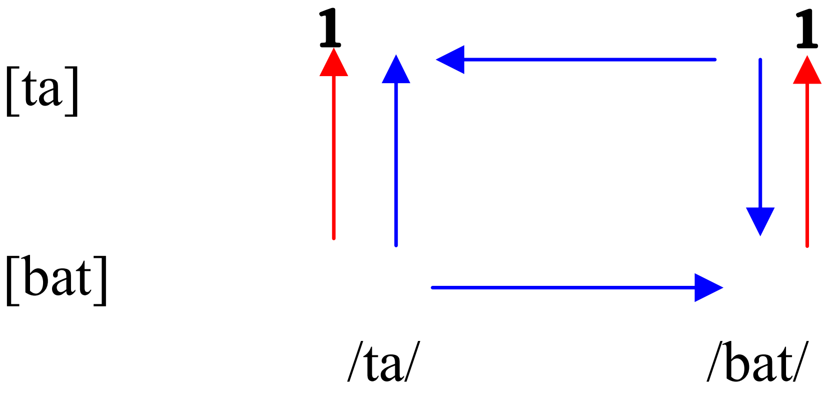
Demonstration of the basic idea

Assume /bat/ as a lexical input & consider two possible surface strings:

.bat. pronounced [bat]

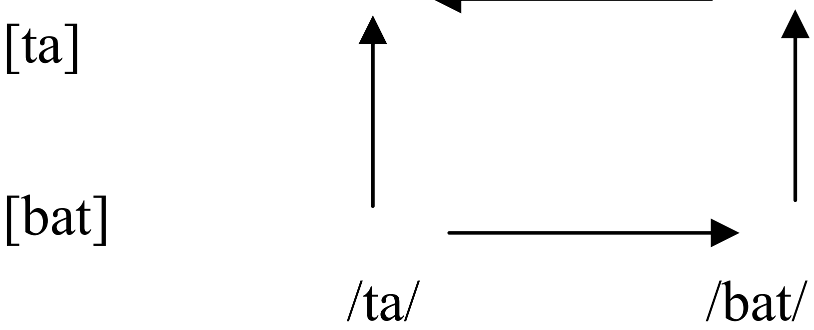
⟨ba⟩.t$\boxed{\text{a}}$. pronounced [ta]

Take /ta/ as another lexical input. As constraints we take FAITH and STRUCTURE, the latter standing for a complex of markedness constraints making [ta] more harmonic than [bat]. Importantly, the same OT Grammar can be uses both for production & comprehension:



Initial State

STRUCTURE(1) ≫ FAITH



resulting

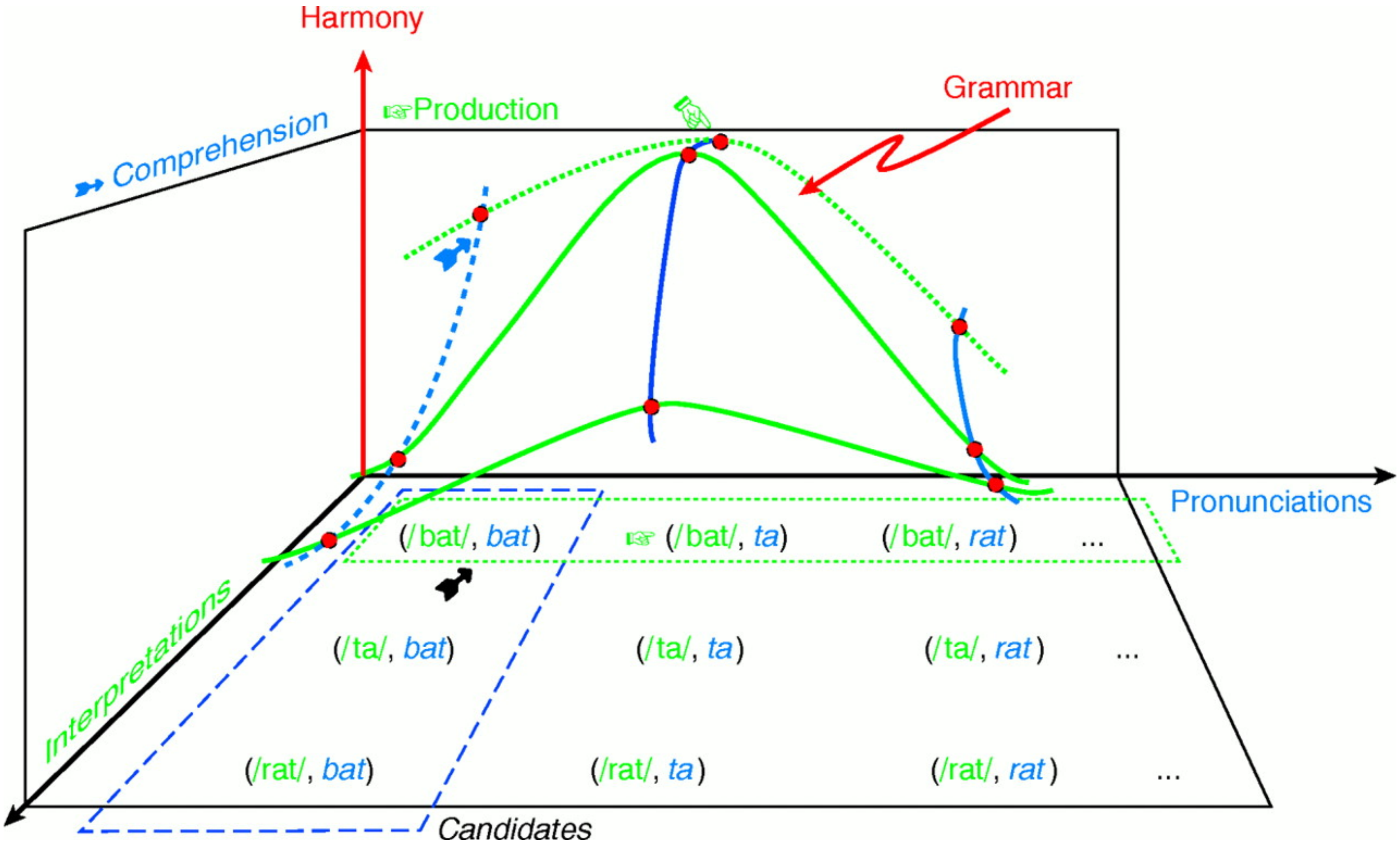
harmonic order

Comprehension: [bat] ⇒ ? Solution /bat/

Production: /bat/ ⇒ ? Solution [ta] **Conflict: associate [bat] !**

In comprehension, [bat] is correctly associated with /bat/. This is a consequence of the fact that the structural (markedness) constraints are sensitive to the overt phonetic forms only. Consequently, FAITH determines the correct association.

On the other hand, **in production** /bat/ is associated (wrongly) with the overt form [ta], which is the most unmarked form. This is a consequence of the fact that within the initial Grammar the faithfulness constraints are dominated by the markedness constraints.



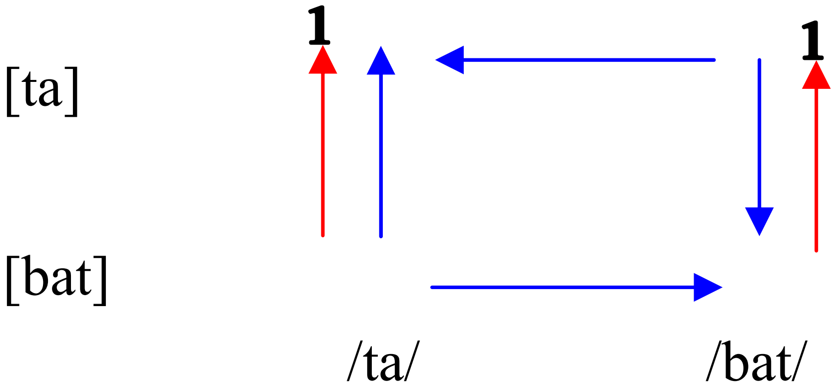
Harmony
☞Production
Comprehension
Grammar
Pronunciations
Interpretations
Candidates
(/bat/, bat)
☞ (/bat/, ta)
(/bat/, rat)
...
(/ta/, bat)
(/ta/, ta)
(/ta/, rat)
...
(/rat/, bat)
(/rat/, ta)
(/rat/, rat)
...

The same idea expressed in a 3D representation (from Prince & Smolensky 1997): The horizontal plane contains pairs such as < /bat/, *ta* > (representing a structure in which the lexical item /bat/ is simplified and pronounced *ta*). The vertical axis shows the relative harmony of each structure, an ordinal rather than a numerical scale. This harmony surface schematically depicts a young child's knowledge of grammar: STRUCTURE dominates FAITHFULNESS.

In **comprehension**, the pronunciation *bat* is given, and competition is between the column of structures containing *bat* (dashed box). Because these are all pronounced *bat*, they tie with respect to STRUCTURE, so lower-ranked FAITHFULNESS determines the maximum-harmony structure to be (/bat/, *bat*), marked with ➻ (peak of the dashed curve). The same grammar that gives correct comprehension results in incorrect—simplified—**production**: the row of structures containing /bat/ compete (dotted box); the maximum-harmony structure best-satisfies top-ranked STRUCTURE with the simplified pronunciation *ta* (peak of the dotted curve): this is marked ☞.

Triggering learning

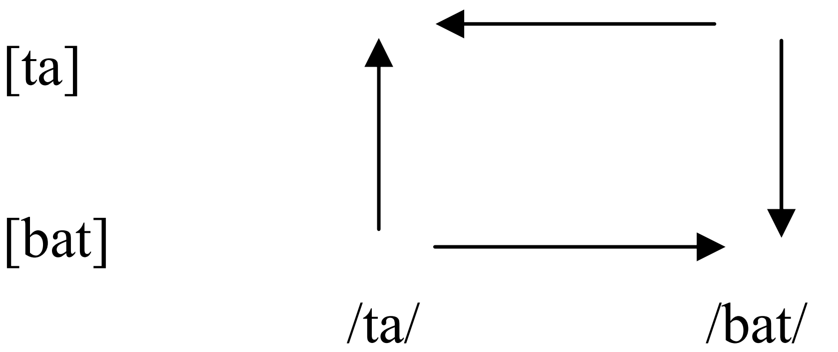
According to Smolensky (1996b) the ‘conflict’ between comprehension and production is the trigger for learning, where learning is understood as a reranking of the involved constraints. In short, the relevant (disturbing) constraints are demoted. In our example, STRUCTURE is demoted:



After learning

FAITH ≫ STRUCTURE(1)

(reranked)



resulting harmonic order

Comprehension: [bat] ⇒ ? Solution /bat/

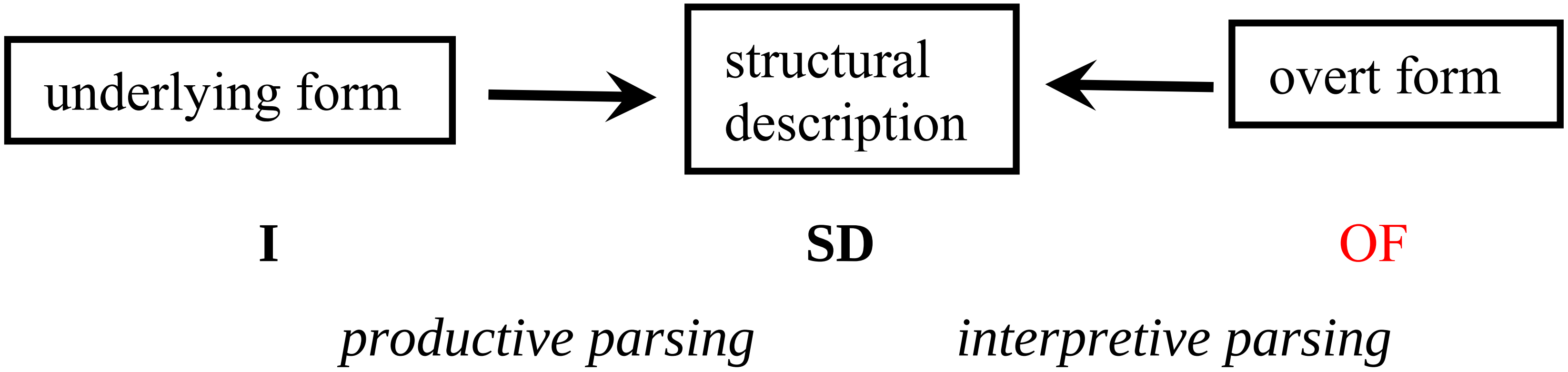
Production: /bat/ ⇒ ? Solution [bat]

4 The OT learning algorithm

- The algorithm starts with an initial grammar: as above, FAITH-fulness constraints are dominated by MARKedness constraints. (This initial ranking is a necessary precondition for a language to be learnable; cf. Smolensky (1996a) for the general argument)

- Comprehension mode: The algorithm proceeds by taking overt phonetic forms as primary data, and assign this data full structural descriptions (*robust interpretive parsing*).

- Production Mode: Determine the current Grammar's output starting with the structural description assigned by the comprehension mode. Since the grammar isn't yet complete this procedure normally doesn't lead back to the origin overt form.



- Constraint Demotion: whenever the structural description which has just been assigned to the overt data (comprehension) is less harmonic than the current grammar's output (production), relevant constraints are demoted minimally to make the comprehension parse the more harmonic.

- This yields a new grammar, which the algorithm then uses to repeat the whole process over again, reassigning structural descriptions to the primary data and then reranking constraints accordingly. The cycle is iterated repeatedly.

- This kind of bootstrap algorithm transforms a bad grammar into a better one. It has been illustrated (simulation) that the algorithm in most cases allows efficient convergence to a correct grammar. (Supposed that the hierarchy of the target language has the property of *total ranking*)

- OT helps to translate structural insights from *Markedness Theory* into a concrete learning algorithm.

- Learning develops a stabilized OT Grammar that can be characterized by the feature of recoverability or bidirectional optimality

- For a critical evaluation of Smolensky & Tesar's (classical) OT learning theory see Hale & Reiss (1998), for an improved learning theory see Boersma & Hayes (2001) [in the reader].

- For a very simple example, see exercise 2. For a couple of more realistic examples and a careful discussion of how the learning algorithm can fail, see chapter 4 of Tesar & Smolensky's (2000) excellent book "Learnability in Optimality Theory". [see a review in the reader]

5 Richness of the base and constraints on inventories

- In standard Generative Grammar, the source of *cross-linguistic variation* is manifold. There are cross-linguistic differences in the input and output systems and in the (parameterised) principles on rules. Especially, the inputs are predominantly determined by language-specific lexical factors.

- OT is a very restrictive theory with regard to the source of variation. Essentially, the following is a fundamental principle of standard OT:

Richness of the Base

The source of all systematic cross-linguistic variation is constraint reranking. In particular, the set of inputs to the grammars of all languages is the same. The grammatical inventories of a language are the outputs which emerge from the grammar when it is fed the universal set of all possible inputs.

(This principle was proposed in Prince and Smolensky 1993:191)

- *Richness of the base* requires that systematic differences in inventories arise from different constraint rankings, not different inputs.
- *Richness of the base* is not a empirical principle but a methodological assumption (rejecting constraints on inputs).

Constraints on inventories

How to explain the different inventories in natural languages? According to OT, the content of lexical inputs is unconstrained. Whether some segment occurs on the surface in a particular language is determined strictly by the constraint grammar of the language in question.

> If faithfulness to a particular feature outranks any prohibitions governing the appearance of the feature, then the feature contributes to defining a language's inventory. If prohibitions against some feature outrank relevant faithfulness constraints, then the feature does not play a role in the inventory

As an example, consider the case of voicing on obstruents. (Recall that obstruents refer to the class of oral stops and fricatives such as the voiceless series p, t, k, s, š and the voiced series b, d, g, z, ž)

- English, German, Dutch, ...: The feature VOICE is **contrastive** for obstruents, i.e. there are minimal pairs like pan/ban, tend/dent, kill/gill, sip/zip mean different things.
- Haiwaiian: The feature VOICE is **noncontrastive** for obstruents. In Haiwaiian, all obstruents (p, k) are voiceless. The voicelessness is redundant.

There is a tendency for obstruents to be voiceless. It derives from the phonetic fact that it is more difficult to maintain vibration of the vocal cords when there is a constriction of the type that produces a fricative or an oral stop.

Phonological markedness constraint

Obstruents must be voiceless: OBS/*VOICE

First case: VOICE a contrastive feature. Voiced obstruents are attested to the inventory if lexical voicing contrasts override this markedness constraint, i.e.

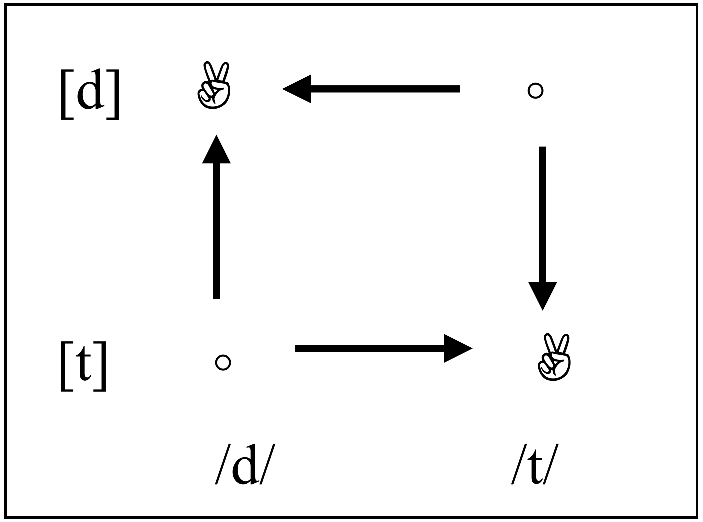
FAITH[VOICE] ≫ OBS/*VOICE.

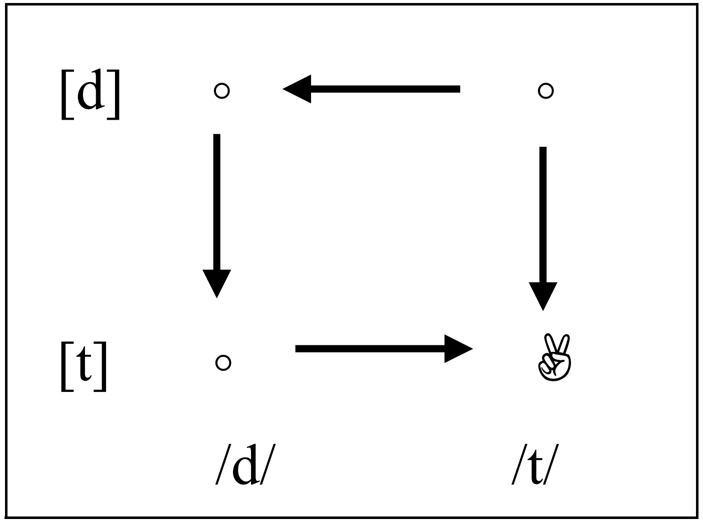
Second case: VOICE as a non-contrastive feature. Voiced obstruents are excluded from the inventory if the markedness constraint overrides faithfulness:

OBS/*VOICE ≫ **FAITH[VOICE]**.

		FAITH[VOICE]	OBS/*VOICE		FAITH[VOICE]	OBS/*VOICE
[d]	☞		*		*	*
[t]		*		☞		
	/d/			/t/		

		OBS/*VOICE	FAITH[VOICE]		OBS/*VOICE	FAITH[VOICE]
[d]		*			*	*
[t]	☞		*	☞		
	/d/			/t/		





[Note: Joan Bresnan (1997, 2001) applies the basic ideas developed in phonology and gives a principled account for a *typology of pronominal systems* and the emergence of the unmarked pronoun.]

Neutralization

In Russian and Dutch voicing is contrastive on obstruents. That is, as in English, the voicing distinction on obstruents leads to differences in lexical meaning (e.g. [bɛ.dən], [bɛ.tən] in Dutch). Unlike English, Russian and Dutch does not maintain the voicing contrast in all positions. Specifically, the distinction between voiced and voiceless obstruents is lost at the end of a syllable where all obstruents appear as voiceless.

As shown earlier, neutralization can described as an extension of the system

FAITH[VOICE] ≫ OBS/*VOICE

by adding the constraint CODA/*VOICE which overrides the other constraints:

CODA/*VOICE ≫ FAITH[VOICE] ≫ OBS/*VOICE

Allophony

As a final example of constraint interaction, a feature may be noncontrastive, but with a distinction nevertheless arising in a predictable context, a case of allophony. In this case, typically a constraint on assimilation overrides the constraints in

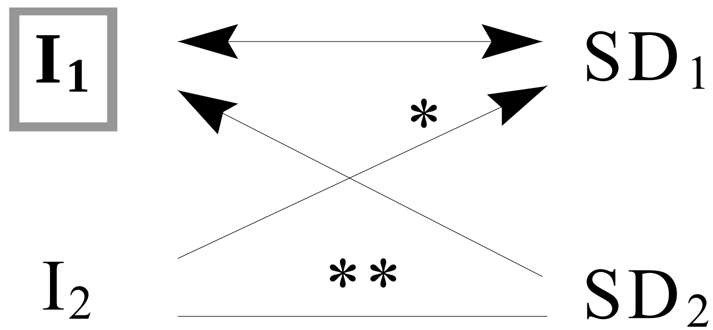
$$\text{OBS/*VOICE} \gg \text{FAITH[VOICE]}$$

(See the example in the exercise part).

Lexicon optimization

The idea is that whenever the learner has no evidence (from surface forms) to postulate a specific divergent lexical form, she will assume that the input is identical to the surface form. Notice that this approach to the analysis of inputs is based on the assumption of full specification and is opposing to the idea of *underspecification* with regard to the inputs.

Lexicon optimization means recoverability of the inputs from the outputs. It invites to introduce a bidirectional mode of optimization.



Only recoverable inputs are assumed to be realized in the mental lexicon.

Lexicon Optimization

Examine the constraint violations incurred by the winning output candidate corresponding to each competing input. The input-output pair which incurs the fewest violations is considered the optimal pair, thereby identifying an input from the output.

(This principle was introduced in Prince and Smolensky (1993) and developed in Itô, Mester & Padgett (1995))