Lecture 5: OT Semantics/Pragmatics

The Aim

Bringing together:

- The tradition of *Radical Pragmatics*
- The view of *Optimality Theory*

Advantages

For *Radical Pragmatics*	For *Optimality Theory*
Improved analyses	New applications
theoretical stringency	Motivating the constraints
the emergence of iconicity	New ideas about grammaticalization and language change

Outline

1. Meaning and Interpretation
2. Blocking and global theories of language
3. Optimality Systems
4. OT and Connectionism
5. The Motivation for Strong Bidirection
6. Radical Pragmatics in OT
7. The Weak Bidirection and Constructional Iconicity
8. Example 1: Negative Strengthening
9. Example 2: Binding Theory

1 Meaning and Interpretation

The observation: Linguistically encoded information doesn't fully specify the truth conditions of a sentence.

- Katz & Fodor (1963): A full account of sentence interpretation has to include more information than that of syntactic structure and lexical meaning.

  a. *Should we take the lion back to the zoo?*
  b. *Should we take the bus back to the zoo?*

- Psycholinguistics: Mental models, situation structure,...
  The tones sounded impure because the hem was torn.

The tones sounded impure because the hem was torn.

SCOTTISH PIPER

Theoretical Models

- **Kaplan's distinction between *character* and *intension***

  intension = character(σ)

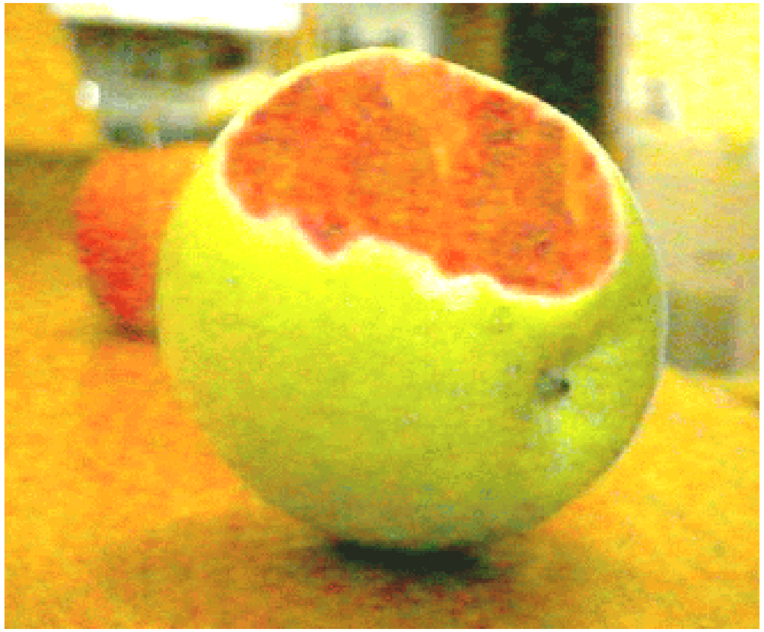
- **Radical Underspecification View**

  Underspecified representations + contextual enrichment (Hobbs 1983, Alshawi 1990, Poesio 1991, Pinkal 1995, etc.)

  => Find optimal enrichments!

- Example for underspecification: Attributive modification

- *a red apple*	[red peel]
- *a sweet apple*	[sweet pulp]
- *a reddish grapefruit*	[reddish pulp]
- *a white room/ a white house*	[inside/outside]



A red apple?

What color is an apple?

Q_1 What color is its peel?

Q_2 What color is its pulp?

2 Blocking and global theories of language

Local Theories

The (grammatical) status of a (linguistic) object LO is decided exclusively considering properties of LO, and the properties of other linguistic objects LO' are completely irrelevant for this decision.

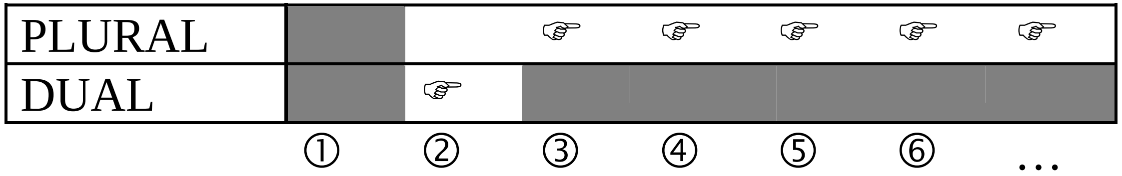
Examples: Traditional Generative Linguistics, Model Theoretic Semantics.

Global Theories (Competition-based)

There are different linguistic objects in competition. The winner of the competition suppresses the other competing candidates, ruling them out from the set of well-formed linguistic objects.

Examples: Early Structuralism (Saussure), Field Theories, Prototype Theories, Optimality Theory, Connectionism.

Blocking



The value of a German or Latin **plural** is not the value of a Sanskrit plural. But the meaning, if you like, is the same. In Sanskrit, there is the dual. Anyone who assigns the same value to the Sanskrit plural as to the Latin plural is mistaken because I cannot use the Sanskrit plural in all the cases where I use the Latin plural.

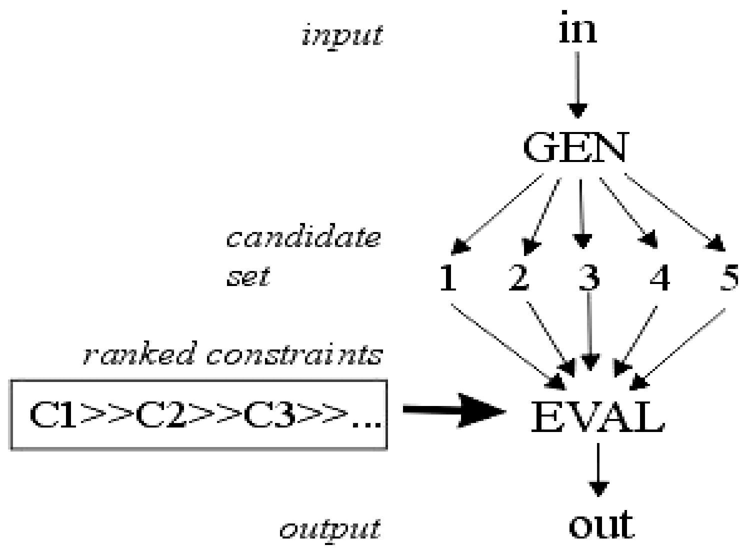
If you take on the other hand a simple lexical fact, any word such as, I suppose, **mouton** (French) may have the same meaning as **sheep** in English. However, it doesn't have the same value. For if you speak of the animal on the hoof and not on the table, you say sheep. It is the presence in the language of a second term (mutton) that limits the value attributable to sheep.

Notes taken by a student of Saussure's lectures [4 July 1911]

3 Optimality Systems

Basics

- An optimality system is an triple ⟨**GEN**, *C*, >>⟩ where
  - **GEN** is a relation — Universal basis
  - *C* is a set of functions from **GEN** to $\mathbb{N}$ — universal constraints
  - $\gg$ is a linear ordering on *C* — language-particular ranking



- The ranking $\gg$ of the constraints constitutes a well-founded preference relation $<$ between pairs π of **GEN** (read $<$ as *less costly* or *more harmonic*):
  $\pi < \pi'$ iff there is a $c \in C$ such that $c(\pi) < c(\pi')$ and for all $c' \gg c$: $c'(\pi) = c(\pi)$

Definition (unidirectional and bi-directional optimality)

Let $\Omega = \langle \mathbf{GEN}, C, >> \rangle$ be an OT-system. Assume that GEN reflects the direction of interpretation; for example with $\langle a, b \rangle \in \mathbf{GEN}$ assume that *a* is a syntactic form and *b* a semantic form.

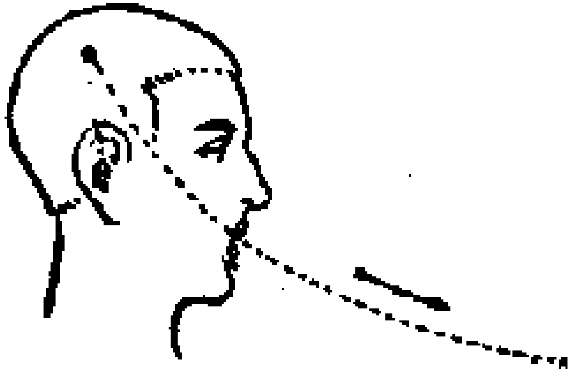
- A pair $\langle a, b \rangle$ is called *Hearer optimal* w.r.t. Ω iff

  (i) $\langle a, b \rangle \in \mathbf{GEN}$
  (ii) there is no *b*' such that $\langle a, b' \rangle \in \mathbf{GEN}$ and $\langle a, b' \rangle < \langle a, b \rangle$

- A pair $\langle a, b \rangle$ is called *Speaker optimal* w.r.t. Ω iff

  (i) $\langle a, b \rangle \in \mathbf{GEN}$
  (ii) there is no *a*' such that $\langle a', b \rangle \in \mathbf{GEN}$ and $\langle a, b' \rangle < \langle a, b \rangle$

- A pair $\langle a, b \rangle$ is called (*strongly*) *optimal* w.r.t. Ω iff it is both Speaker and Hearer optimal.



Speaker

Phonology, Morphology:
Prince & Smolensky (1989);
McCarthy & Prince (1993); …

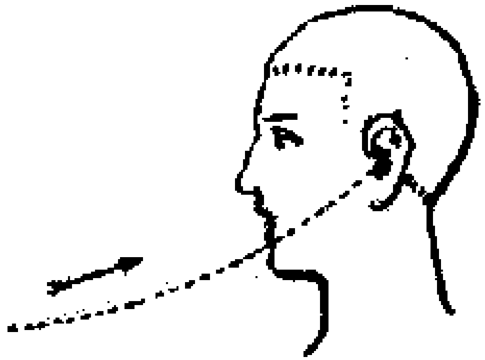
Syntax: Grimshaw (1997);
Bresnan (1999); …

Optimal Generation

Semantics: de Hoop & de Swart (1999) ; de Hoop & Hendriks (2001)

E.g. Domain Restrictions:

- *Most linguists sleep at night*
- *Most linguists drink at night*



Hearer

Optimal Interpretation

4 Optimality Theory and Connectionism

- According to Smolensky, OT has proposed a new computational architecture for cognition which integrates connectionist and symbolic computation.
- Integrative connectionism: Seeing Symbolism as a way to describe higher order properties of neuronal networks.
- Reductionism as a very powerful *research instrument* for justifying higher order regularities
- Which properties can be reduced and which not?
  - Optimality (the correct output maximizes H)
  - Particular systems of constraints (evolutionary grounding)
  - Strictness of domination (automatic processing)
  - Bidirection (learning, evolutionary grounding)
  - Universality
  - Factorial Typology

Hopfield networks

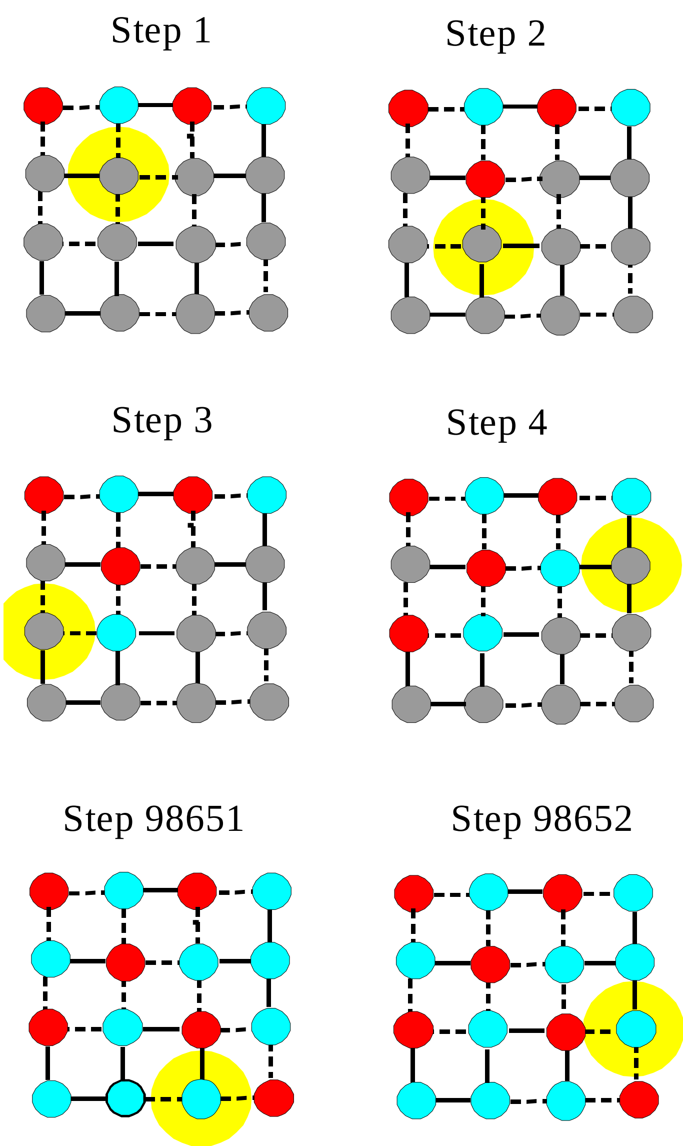
Let the interval [-1,+1] be the *working range* of each neuron

+1: maximal firing rate
0: resting
-1 : minimal firing rate)

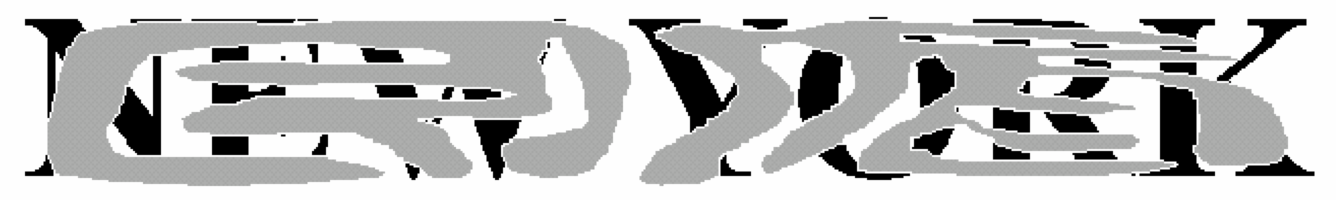
$S = [-1, 1]^n$
$w_{ij} = w_{ji}$, $w_{ii} = 0$

ASYNCHRONOUS UPDATING:
$s_i(t+1) = \theta(\Sigma_j w_{ij} \cdot s_j(t)$, if $i = \text{random}(1,n)$
$s_i(t+1) = s_i(t)$, otherwise

(θ nonlinear function: threshold)

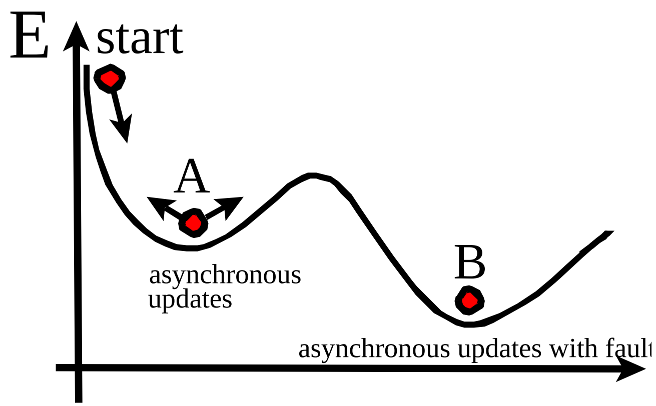


Pattern Completion



Pattern Completion and Bidirection

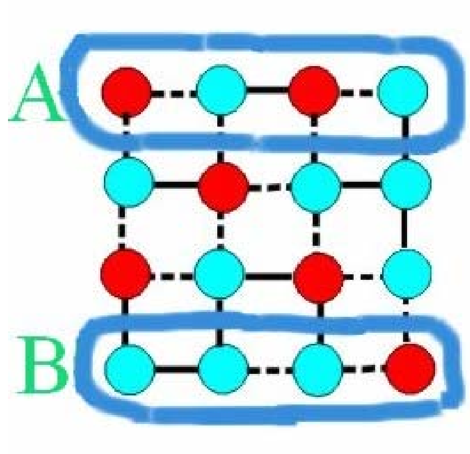
- The stabilized results of asynchronous updating (with faults; adiabatic freezing) can be described as the local (global) minima of the Ljapunov/energy function E (clamping the given input)



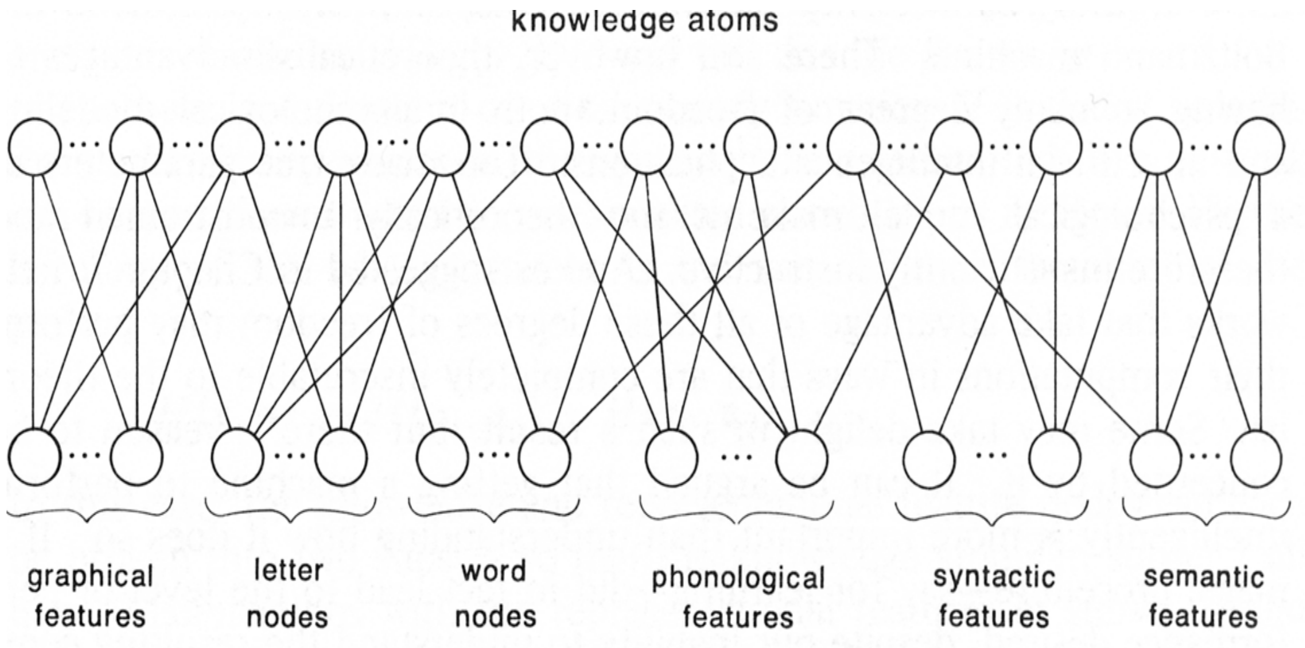
- One and the same pattern can be produced starting from different inputs:

  $S_A \rightarrow X \leftarrow S_B$

  However, there is no warranty of bidirection: If S_A produced a pattern X that contains S_B, then starting with S_B does not necessarily result in X (with S_A as part)



Harmony Theory



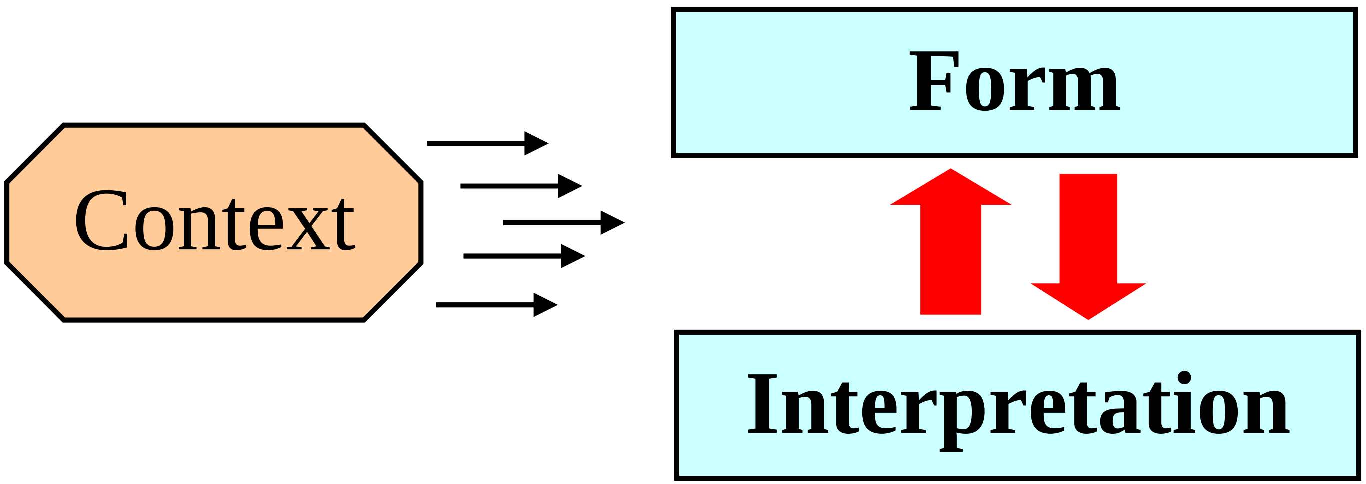
The same holds for Harmony Theory: Harmony theory does **not** say that the different optimisations converge when we start with different parts of a lucid representational vector. The theory says only that one and the same Ljapunow function (=system of ranked constraints in OT) can be used when the system operates like a hearer (starting with a natural language form and ending with an interpretation) or when it operates like a speaker (starting with an activated interpretation and ending with a form).

The asymmetry

Hence, the theory does **not** say that we come back to the original expression when we execute both operations in succession. And that seems empirically right.

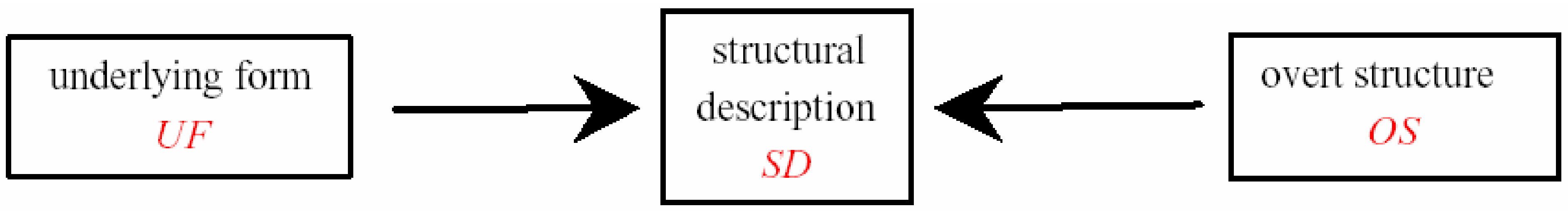
- The phenomenon of aphasia illustrates possible asymmetries in production and comprehension (e.g. Jakobson 1941/1962).

- A related asymmetry is found in language acquisition. It is well known that children‘s ability in production lags dramatically behind their ability in comprehension.



5 The Motivation for Strong Bidirection

In overcoming the lag between production and comprehension, a kind of bootstrap mechanism seems to apply that makes crucially use of the *robustness* of comprehension, an issue that is substantial for the OT learning theory (Smolensky 1996, Tesar & Smolensky 2000).



productive parsing (*production mode*)

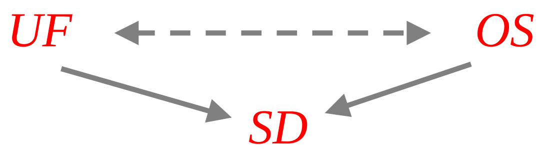
interpretive parsing (*comprehension mode*)

The discrepancy between interpretive parsing and productive parsing triggers learning.

- After learning, the two modes of assigning structure to inputs, productive and interpretive parsing, coincide.

Tesar's and Smolensky's learning theory

T&S's paradigm is *auto-associative learning*: Presentation of overt stimuli => the system is supposed to extract the structure of the input patterns. Examples from phonology, syntax.



1. The learner assigns to *OS* an optimal pair <*UF*, *SD*> (interpretive parsing).

2. Applying production directed optimization to *UF* resulting in *SD'*.

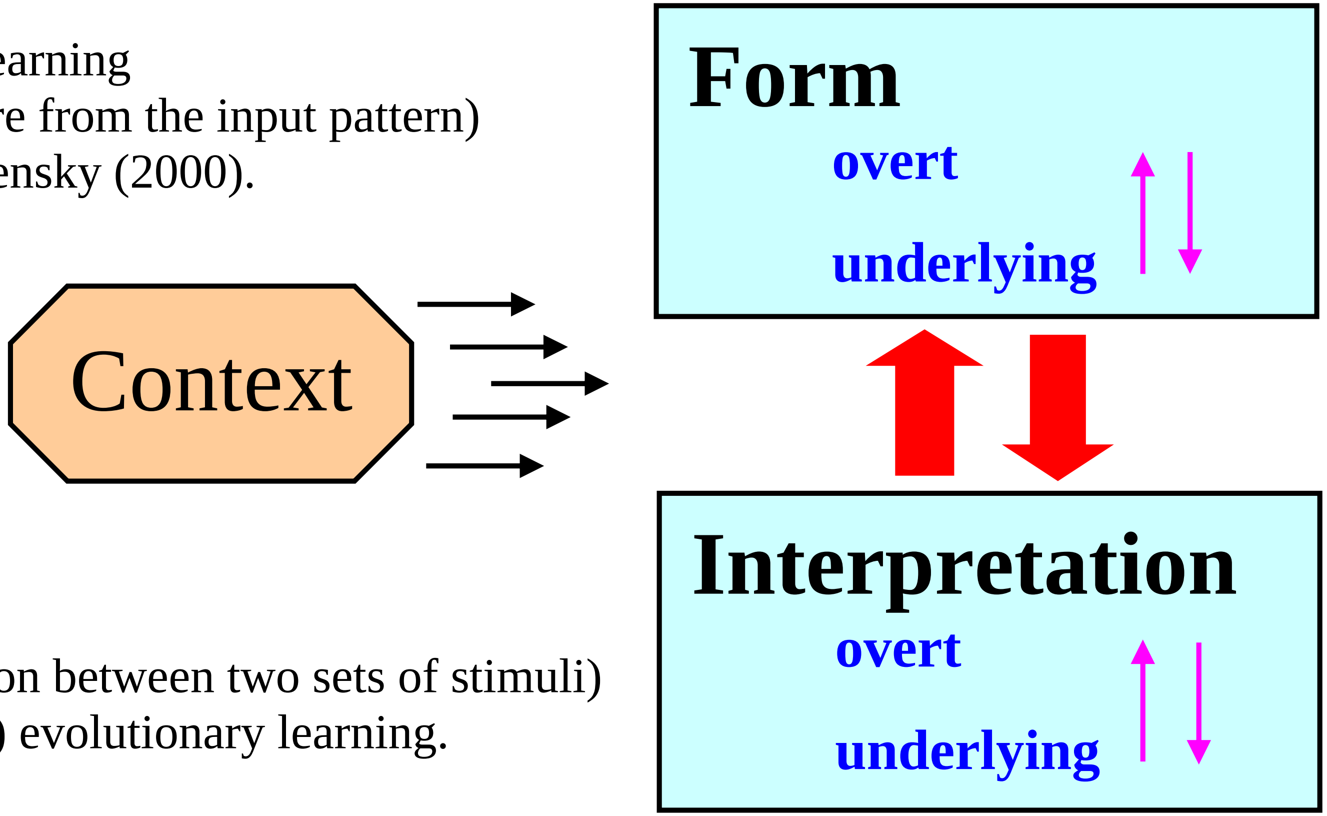
3. If *SD* = *SD'* then ***H*** doesn't need adjustment.

4. If *SD* $\neq$ *SD'*, then ***H*** needs adjustment. The learner assumes *SD* to be corrects and applies constraint demotion with *SD* as winner and *SD'* as loser. (Constraints that are violated (more often) by *SD* are reranked below constraints that are violated by *SD'*).

The result of stabilization is SYMMETRY: if *OS* =>*UF*, then *UF* => *OS*. As a consequence, all hearer-optimal pairs are strongly optimal!

Two kinds of learning

(A) Auto-associative learning
(extracting structure from the input pattern)
e.g. Tesar & Smolensky (2000).



(B) Pattern association
(learning the relation between two sets of stimuli)
e.g. Jäger's (2002) evolutionary learning.

Pattern association

- A set of pairs of patterns are repeatedly presented. The system is to learn that when one member of the pair is presented it is supposed to produce the other. In this paradigm one seeks a mechanism in which an essentially arbitrary set of input patterns can be paired with an arbitrary set of output patterns.
- For example, input patterns can be lexigrams (e.g. senseless syllables), and output patterns can be pictures of fruits. Assume a 1-1 correspondence between syllables and pictures.
- If subjects are qualified to match Stimulus A to B and then, without further training, match B to A, they have passed a test of symmetry.
- Children as young as 2 years pass the symmetry test! (Green 1990). Hence, bidirectionality seems to build in the basic learning mechanism.

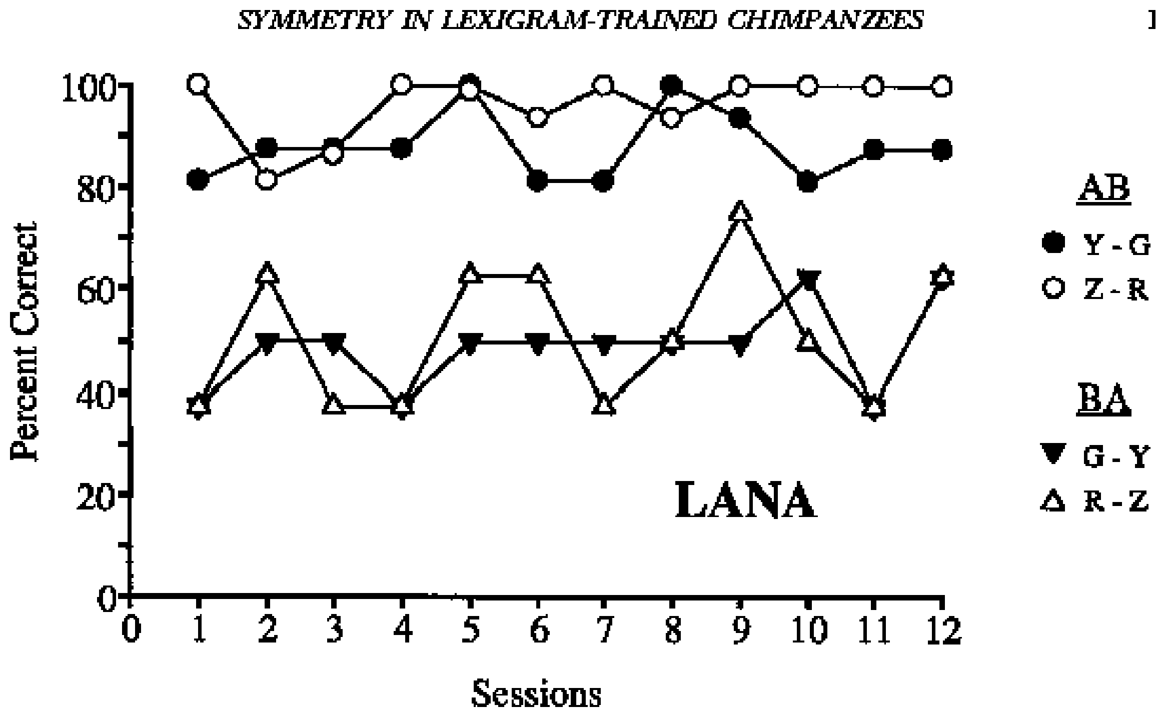
Again, the result is SYMMETRY: If a => b then b => a, and vice versa. As a consequence, all hearer-optimal pairs are strongly optimal! The same for Speaker-optimal pairs.

Washoe in conversation with Roger Fouts (ASL). W. taught some of her human signs to an adopted baby, which is a very interesting case of "cultural transmission".

Lana has learned to speak by punching logos on a computer keyboard. When she pressed a key, the symbol light up. By going so she asked her trainer for an orange.

Chimps and Bidirection

Having learned to match lexigram comparisons to object samples, the chimps were not able, without further training, to match the same objects now presented as comparisons to the corresponding lexigrams, now presented as samples (i.e., they did not show symmetry; see Savage-Rumbaugh, 1984).



From Dugdale & Lowe (2000), p. 15

Kanzi - a Monobo Monkey

Sue Savage-Rumbaugh was trying to teach Kanzi's mom, Matata, a symbolic language.

Kanzi sat on her lap during these sessions. And while Matata did poorly, Kanzi learned.

Kanzi's knowledge was reciprocal. There was no need taught her separately to produce and to comprehend.

6 Radical Pragmatics in OT

"Radical pragmatics is the hypothesis that many linguistic phenomena which had previously been viewed as belonging to the semantic subsystem, in fact belong to the pragmatic subsystem."

Preface to *Radical Pragmatics* (Peter Cole, ed., Academic Press 1981)

- Division of labor between semantics and pragmatics
  - discriminating meaning and interpretation
  - the idea of underspecification

- Gricean mechanism of interpretation
  - Neo-Gricean account (Atlas & Levinson 1981, Horn 1985, ...)
  - Relevance Theory (Sperber & Wilson 1986, Carston 1998, ...)

Conversational Implicatures: Some Standard Examples

(Q1) Some of the boys are at the party
=> Not all of the boys are at the party
(*Scalar implicatures*, Gazdar 1979)

(Q2) Rick is a philosopher or a poet
=> Rick is not both a philosopher and a poet
(*Scalar implicatures*, Grice 1968; Atlas and Levinson 1981)

(Q3) Rick is a philosopher or a poet
=> Rick may (not) be a philosopher; Rick may (not) be a poet
(*Clausal implicatures*, Gazdar 1979; Atlas and Levinson 1981)

(I1) If you mow the lawn, I'll give you $5
=> If and only if you move the lawn, will I give you $5
(*Conditional perfection*, Geis & Zwicky, 1971)

(I2) John unpacked the picnic. The beer was warm.
=> The beer was part of the picnic.
(*Bridging*, Clark & Haviland, 1977)

(I3) John said 'Hello' to the secretary and then he smiled
=> John said 'Hello' to the female secretary and then he smiled
(*Inference to stereotype*, Atlas & Levinson 1981)

I-principle (termed R by Horn)	Q-principle
Quantity 2, Relation	Quantity 1
Say no more than you must (given Q) (Horn 1984)	*Say as much as you can (given I)* (Horn 1984).
Read as much into an utterance as is consistent with what you know about the world (bearing the Q-principle in mind). [Levinson 1983: 146f.]	*Do not provide a statement that is informationally weaker than your knowledge of the world allows, unless providing a stronger statement would contravene the I-principle* [Levinson 1987: 401]
Conditional perfection, *neg-raising*, bridging	Scalar implicatures
Seeks to select the most *harmonic* interpretation	Can be considered as a blocking mechanism
Interpretive Optimization	**Expressive Optimization**

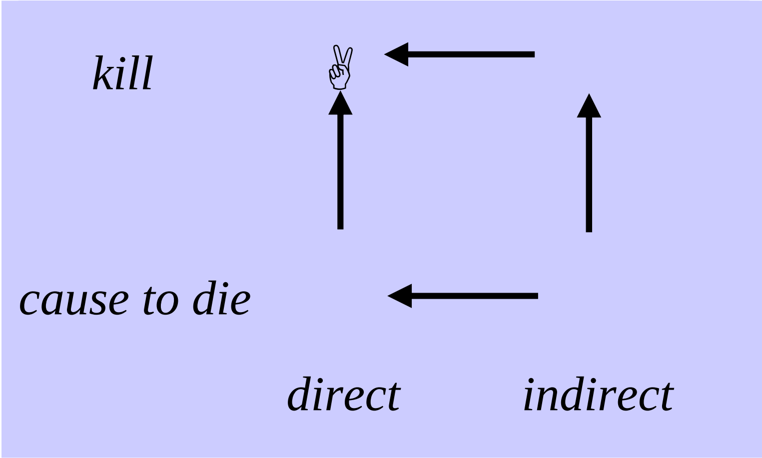
Example: Total Blocking

- **GEN** $\subseteq$ **FORM** x **INT** (Semantics with Underspecifation)

- Markedness constraints for forms and interpretations
  - $\langle \textit{form}, \textit{int}\rangle < \langle \textit{form}', \textit{int}\rangle$ iff *form* is lighter than *form*'
  - $\langle \textit{form}, \textit{int}\rangle < \langle \textit{form}, \textit{int}'\rangle$ iff *int* is more salient than int'

- McCawley's pair:

  Bill killed the Sheriff
  Bill caused the Sheriff to die



- The solution concept of *strong optimality* accounts for total blocking.
  It does not account for partial blocking! Look for other solution concepts!!

7 Weak Bidirection and Iconicity

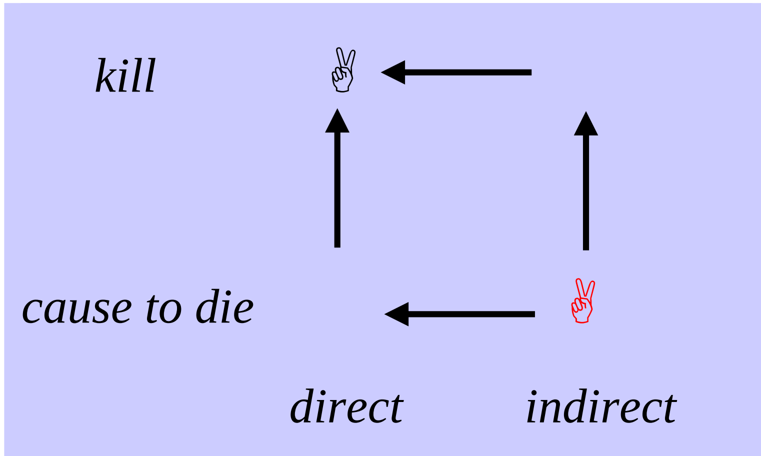
Strong Bidirection does not account for *partial blocking*. Fortunately, there is another conception of bi-directional optimization, a conception that makes use of recursion.

Let $\Omega = \langle$**GEN**, C, >>$\rangle$ be an OT-system. Then a pair $\langle a, b\rangle$ is super-optimal w.r.t. Ω iff

(1) $\langle a, b\rangle \in$ **GEN**
(2) there is no super-optimal $\langle a, b'\rangle < \langle a, b\rangle$
(3) there is no super-optimal $\langle a', b\rangle < \langle a, b\rangle$

John McCawley's example again:

Bill killed the Sheriff
Bill caused the Sheriff to die



Krifka's example: How much precision is enough?

Krifka's Observation

- Vague interpretations of measure expressions are preferred if they are short
- Precise interpretations of measure expressions are preferred if they are long

A: The distance between Amsterdam and Vienna is one thousand kilometers.
B: #No, you're wrong, it's nine hundred sixty-five kilometers.

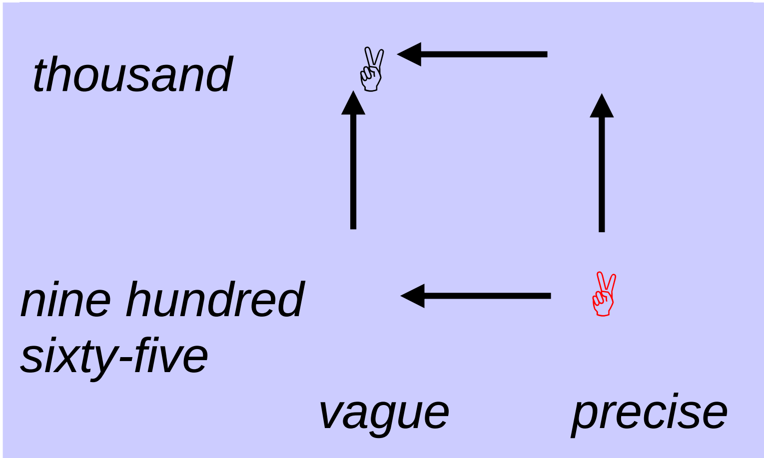
A: The distance between A and V is nine hundred seventy-two kilometers.
B: No, you're wrong, it's nine hundred sixty-five kilometers.

Street sign in Kloten, Switzerland.

Explanation

- Markedness constraints for forms and interpretations
  - ⟨*form*, *int*⟩ < ⟨*form*', *int*⟩ iff *form* is lighter than *form*'
  - ⟨*form*, *int*⟩ < ⟨*form*, *int*'⟩ iff *int* is less precise than int'

- Weak Bidirection

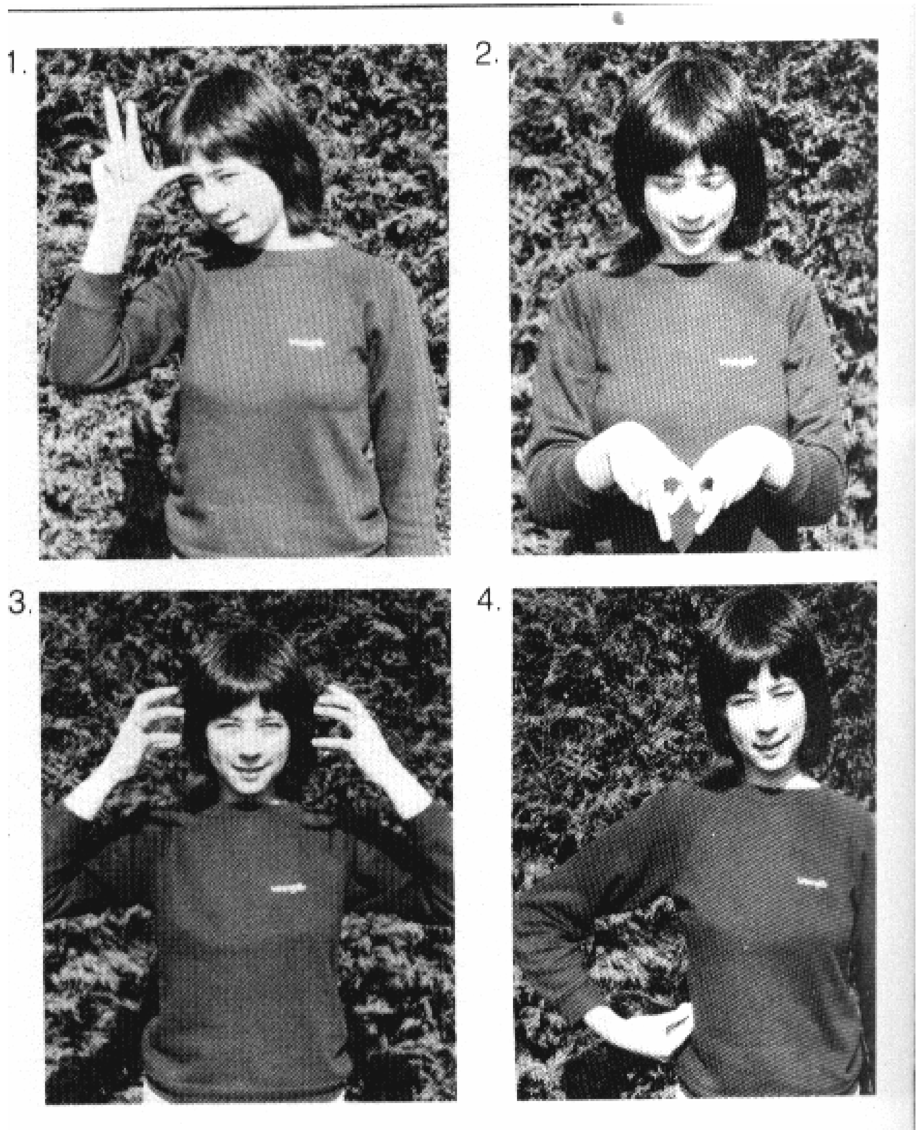


➔ Generalization: Constructional Iconicity in Natural Language

Iconic Signs

Iconic signs reflect part of the external world the sign is intended to designate.

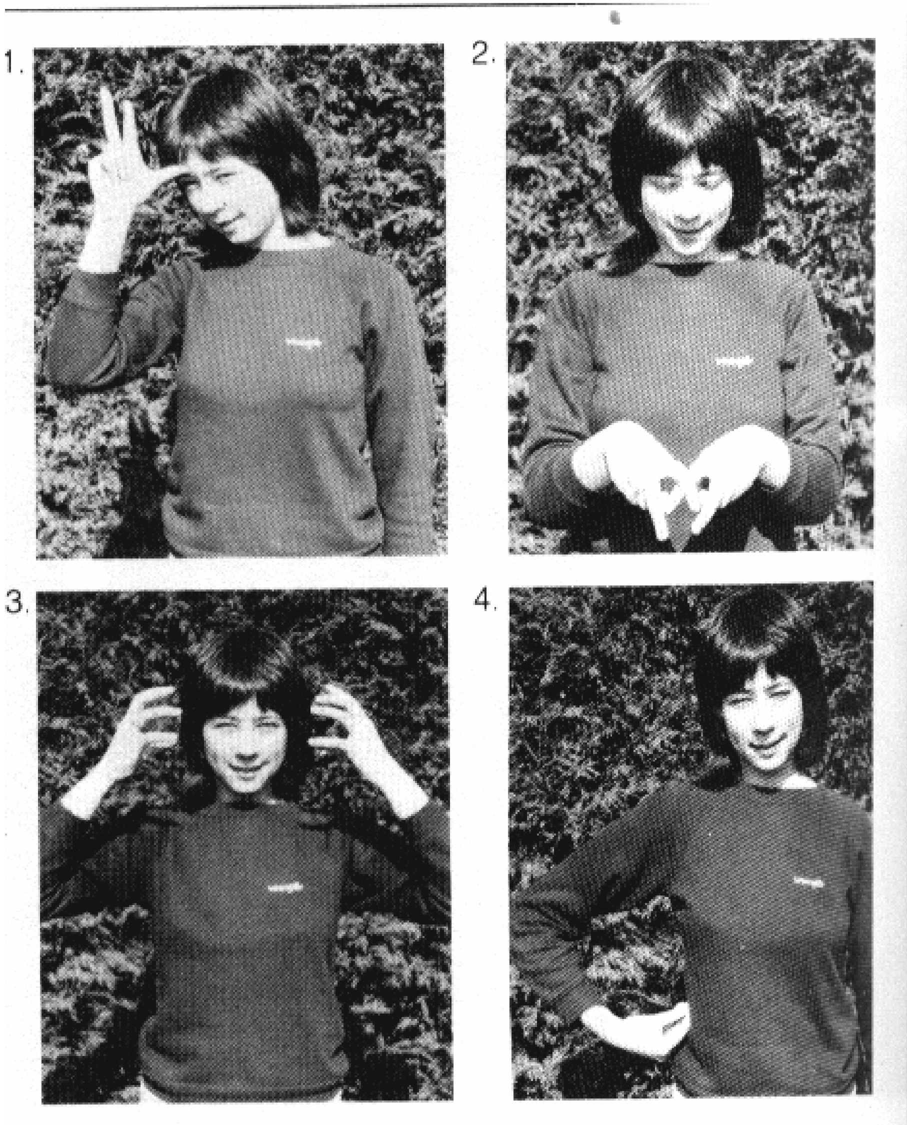
1.
2.
3.
4.



[from David Crystal, Cambridge encyclopaedia of language]

Iconic Signs

Iconic signs reflect part of the external world the sign is intended to designate.



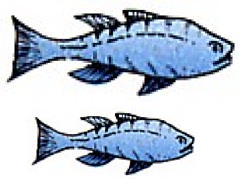
1. Chicken
2. Horse
3. Lion
4. Monkey

[from David Crystal, Cambridge encyclopaedia of language]

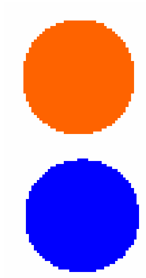
Constructional Iconicity (or Horn's division of pragmatic labor)

Unmarked forms tend to be used for unmarked situations and marked forms for marked situations. (Levinson's M-principle)

- MAYERTHALER ZICK ZACK



- BERLIN & KAY MOLA MILI



- ARGUMENT LINKING (Uszkoreit, Bresnan, Jackendoff, Kiparski, ···)

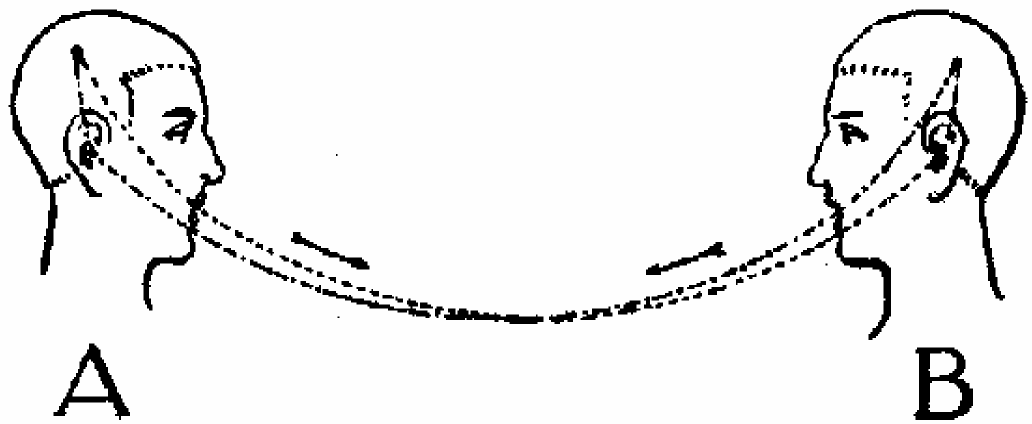
  Agent > Instrument > Recipient/Experiencer > Theme > Location
  Subject > Object$_d$ > Object$_i$ > Oblique

Harmonic alignment

Economy and Language

(I) Economy plays a crucial role in online interpretation and production (e.g. in explaining garden path effects). (Standard OT, Levinson)

(II) Economy constitutes *languages* as conventional systems. (Horn, Zipf)



Georg K. Zipf (1949)

Human Behavior and the Principle of Least Effort. Addison-Wesley. Cambridge 1949.

Two basic and competing forces

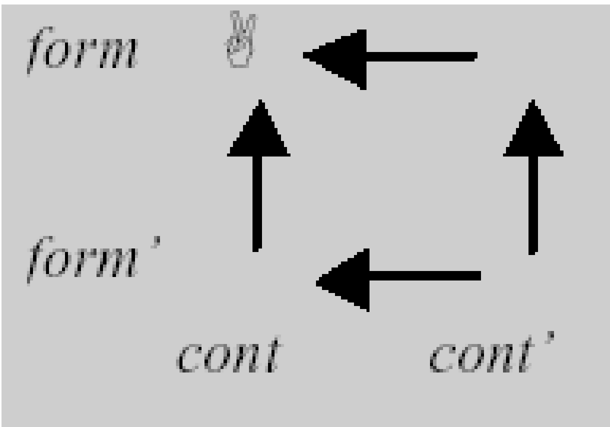
Speaker's economy	Hearer's economy
Force of unification	Force of diversification

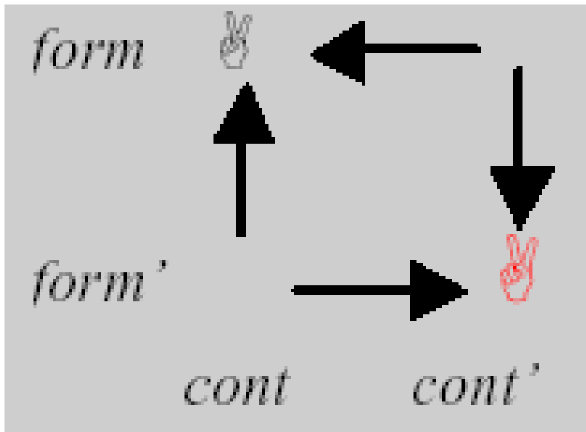
- The two opposing economies are evolutionary forces
- They are balanced during language evolution.

Why two conceptions of Bidirection?

- Strong Optimality as a synchronic law (describing an equilibrium that results from successful learning)

- Weak (Super-) Optimality as a diachronic law (describing the probable outcomes of language evolution under highly idealized conditions)

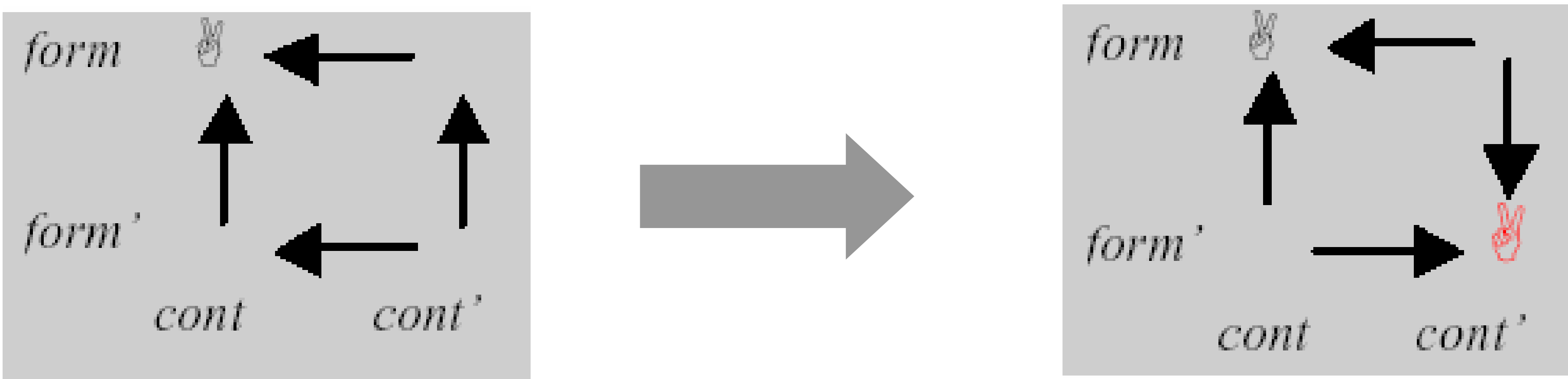




Calculating super-optimal solutions

Jäger (1999), Dekker & van Rooy (1999), Beaver (in preparation) have proposed procedures that update preferences in OT systems such that

(i) optimal pairs are preserved
(ii) a new optimal pair is produced if and only if the same pair was super-optimal at earlier stages.



The evolutionary grounding of weak bidirection

There are many different ways to realize a evolutionary perspective. Different versions highlight the role of *correlations*, *learning*, *mutations*, and the *initial state*, respectively.

- Van Rooy (2002): *Signalling games and evolutionary stable Horn-strategies.*
- Jäger (2002): *Learning constraint sub-hierarchies. The Bidirectional Gradual Learning Algorithm.*
- Blutner, Borra, Lentz, Obdeijn, Uijlings, and Zevenhuijzen (2002): *Signalling Games: hoe evolutie optimale strategieën selecteert.*

Basic Ideas

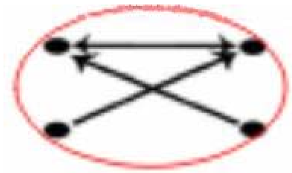
- Each agent is described by an OT-system $\Omega = \langle \mathbf{GEN}, C, >> \rangle$. Within the population Gen and C are fixed, >> may vary.
  Each agent X determines a *speaker's strategy* S_X : Contents => Forms
  and a *hearer's strategy* H_X : Forms => Contents

- In pairwise interactions between an agent *a* (in the role of the speaker) and an agent *b* (in the role of the Hearer) an utility/fittness function U is realized:
  $U(a,b) = \Sigma\, P(i)\, [\, \delta(H_b(S_a(i)), i) - k(S_a(i))],$
  where $\delta(x,y) = 1$ if $x = y$, 0 elswhere. P(i) probability of "content" i, k(f) cost of signal f.

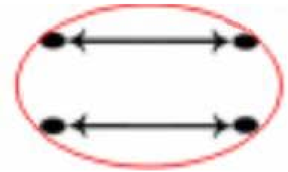
- The agents of the population randomly encounter one another in pairwise interaction. Each organism plays only one, but leaves its offspring behind, where the number of offspring is determined by the utility value U(a, b). Mutations change the strategies played by some elements of the population. After many plays of the game, a strategy yielding a higher number of expected offspring will gradually come to be used by larger and larger fractions of the population.

OT systems and strategies

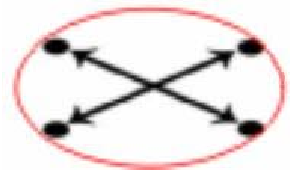
- Smolensky: { ***f*← , *c*← }** >> **…**



- Horn: { ***f*** ← ***c*** , ***c*** ← ***f*** } >> …



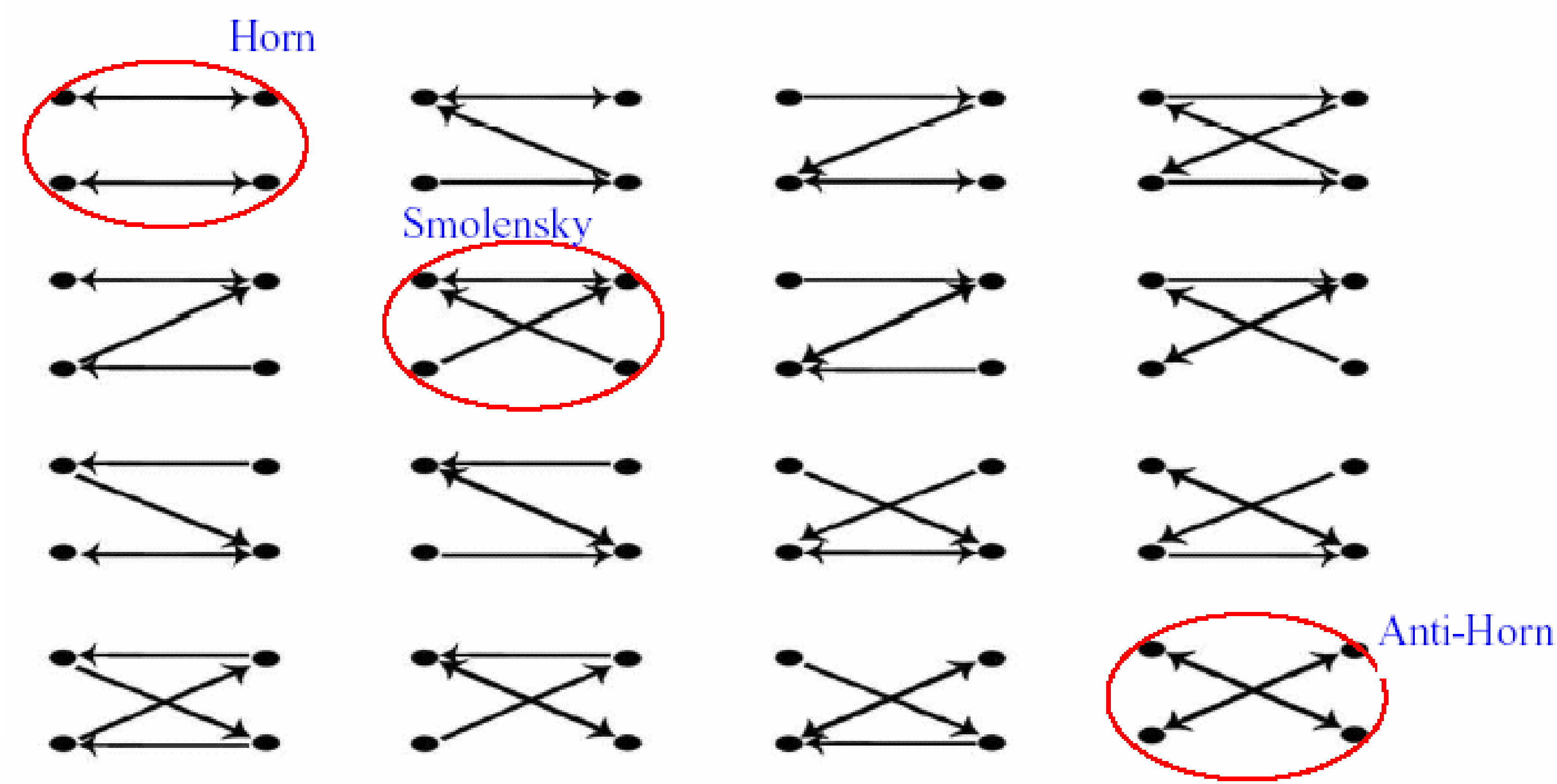
- Anti-Horn: { ****f*** ← ***c*** , ***c*** ← ****f*** } >> …



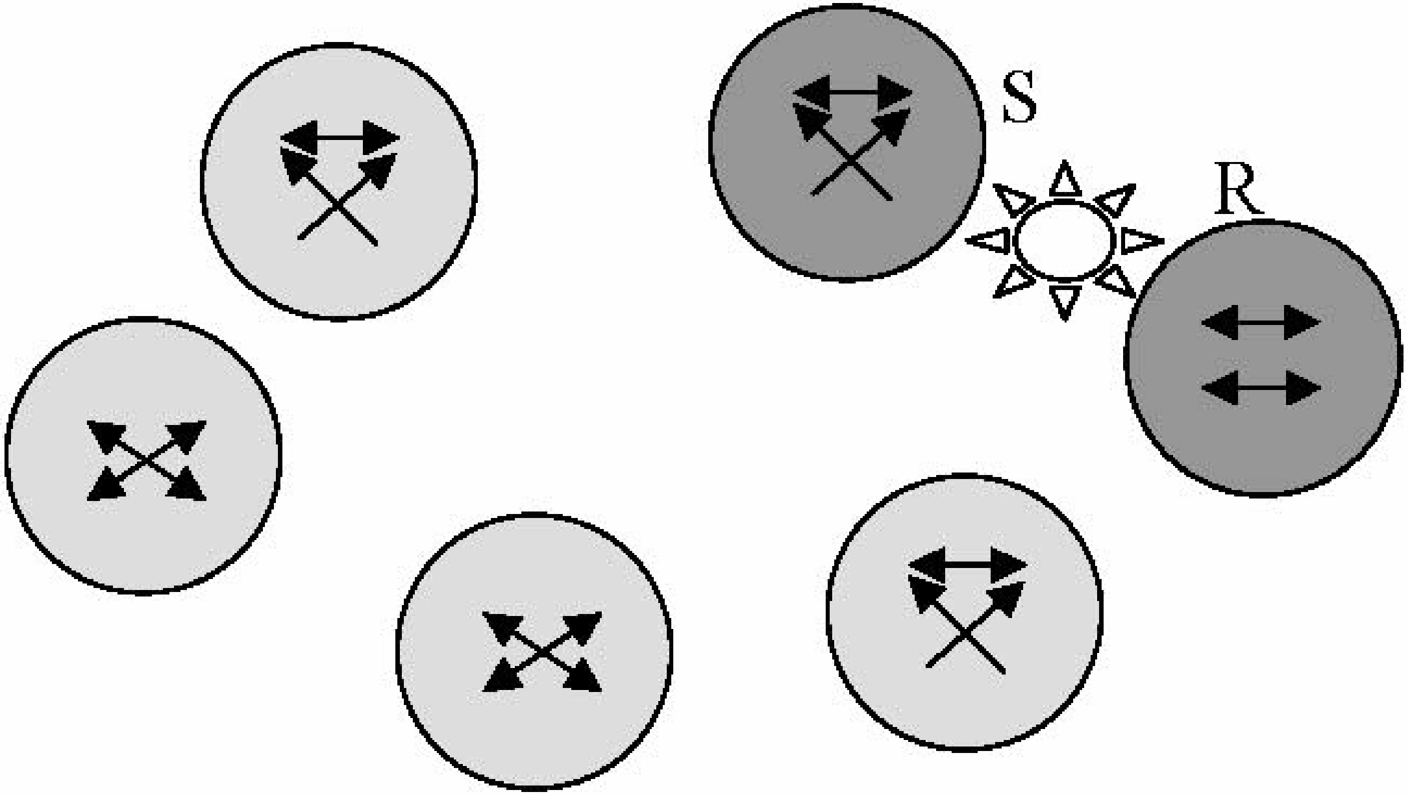
	***f*←**	***c*←**	***f* ← *c***	****f* ← *c***	***c* ← *f***	***c* ← *f**
⟨*f*, *c*⟩		*		*		
⟨*f*, *c*'⟩		*			*	
⟨*f*', *c*⟩	*		*			
⟨*f*', *c*'⟩	*					*

All possible strategies

for an OT-system with GEN = {⟨*f*, *c*⟩, ⟨*f*, *c*'⟩, ⟨*f*’, *c*⟩, ⟨*f*’, *c*'⟩}

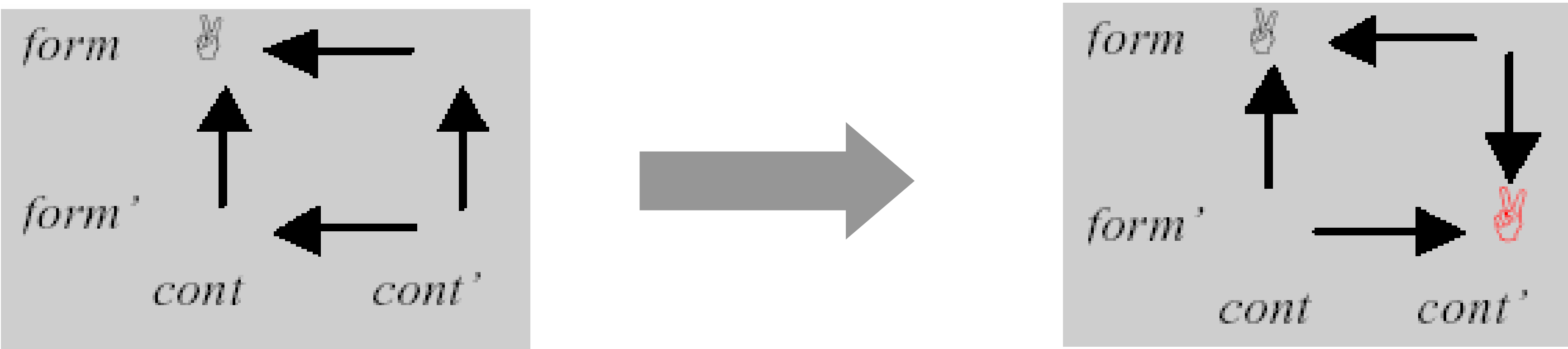


Population and pairwise interaction



Results

- Horn and Anti-Horn are the only strategies (OT-systems) that are evolutionary stable
- Starting with a uniform *Smolensky* population will always result in a pure *Horn* population supposed P(*c*) > P(*c'*) and k(*f*) < k(*f'*)



- Mixed populations develop into pure Horn populations (supposed P(*c*) > P(*c'*) and k(*f*) < k(*f'*))

8 Example : Negative strengthening

What are the effects of **negating** gradable adjectives?

(1) I'm not happy ☺ 😐 ☹

a. *Entailment:* It isn't the case that I'm happy

b. *Implicature*: I'm unhappy

c. *defeasibility*: I'm not happy and not unhappy

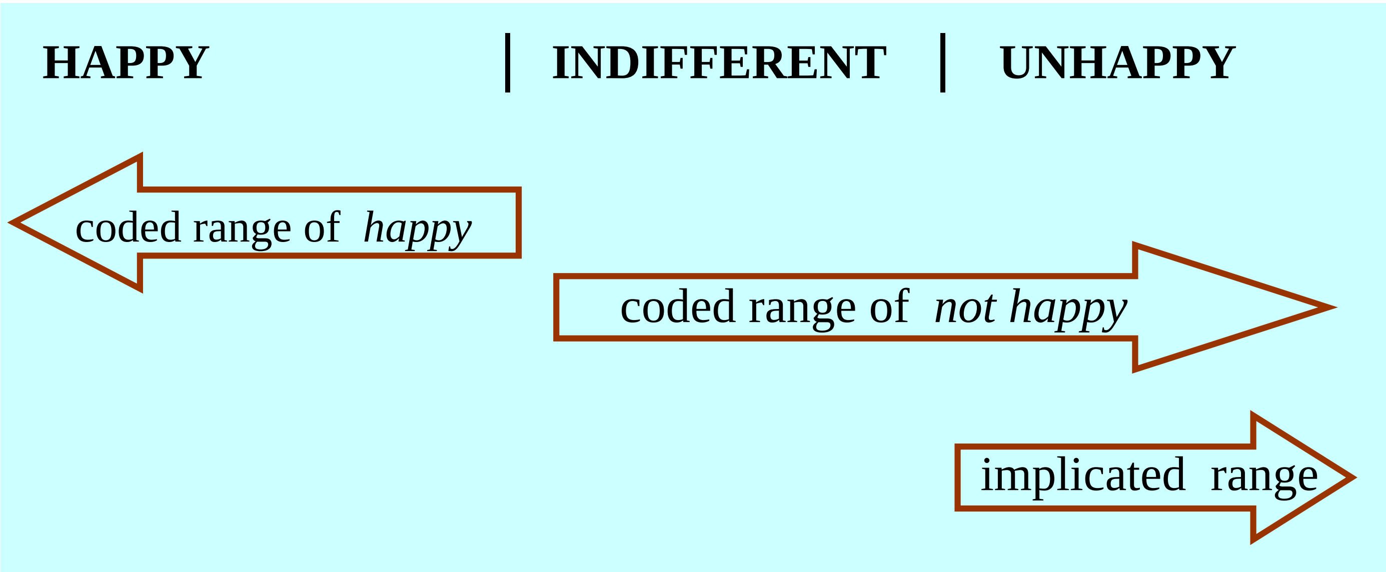


Fig.1 Contradictories implicating contraries

The described effect of strengthening is restricted to the positive (unmarked) elements of antonym pairs!

Litotes

(2) I'm not unhappy ☺ 😐 ☹

a. *Entailment:* It isn't the case that I'm unhappy

b. *Implicature*: I'm rather happy (but not quite as happy as using the expression *"happy"* would suggest)

c. *defeasibility*: I'm not unhappy, in fact I'm happy

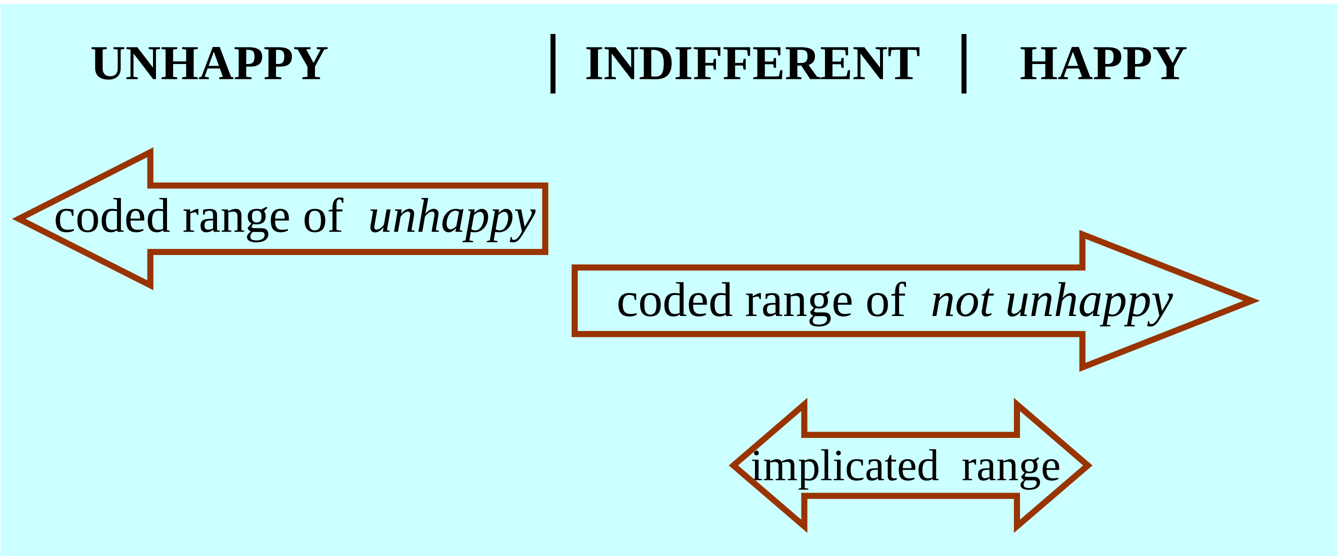


Figure 2: Litotes: when two negatives don't make a positive

Theoretical Assumptions

- The coded range of form-interpretation pairs is due to a three-valued logic: *not* corresponds to weak negation and *un*- to strong negation.

- The number of the involved negation morphemes determine the markedness of the forms

$$\langle \mathit{form}, \mathit{int} \rangle < \langle \mathit{form}', \mathit{int} \rangle \text{ iff}$$

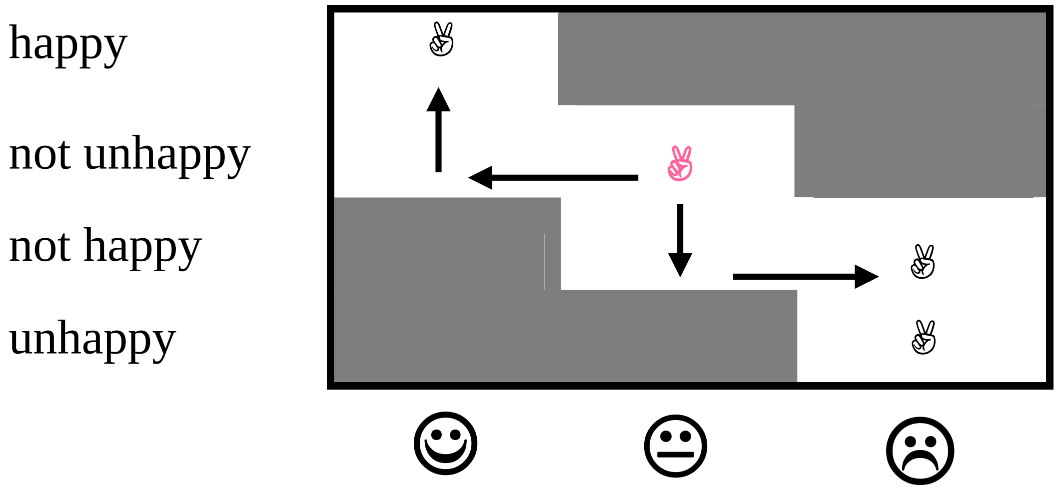
form contains less negation morphemes than *form*'

- The markedness of interpretations decreases towards the ends of the scale (and is maximum in the "neutral" middle)

$$\langle \mathit{form}, \mathit{int} \rangle < \langle \mathit{form}, \mathit{int}' \rangle \text{ iff}$$

int is closer to the end of the scale than *int'*

Super-optimal pairs



9 Example: Binding Theory

(1) Tom$_i$ disagrees with himself$_\mathbf{j}$ *coref* $_{i=j}$ **disj* $_{i\neq j}$

(2) Tom$_i$ disagrees with him$_\mathbf{j}$ **coref* $_{i=j}$ *disj* $_{i\neq j}$

Classical account as a *local theory* (Chomsky)

Principle A: A reflexive must be bound locally *coref* ← *refl*

Principle B: A pronominal must be free locally *disj* ← *pron*

The complementary distributions of reflexives and pronouns follows from complementary constraints

Burzio's account as a *global theory*

PRINC A: A reflexive must be bound locally *coref* ← *refl*

REFEC: In its binding domain (coreference under c-command) a reflexive element is preferable to a pronoun *refl* ← *coref*

PRINC A >> REFEC (look at disjoint interpretation)

The complementary distributions of reflexives and pronouns follows from the fact that pronouns just fill the space from which reflexives are excluded.

Russian

Ja emu$_i$ *skazal vse o *sebe*$_i$ */ nem*$_i$
*I him told all about *Refl*$_i$ */ him*$_i$
'I told him$_i$ everything about himself$_i$'

Levinson’s A-first account

Princ A

PRINC A : *coref* ← *refl* (lexical stipulation)

RefEc

M : *disj* ← *pron*

	coref	*disj*
refl	✌	
pron		✌

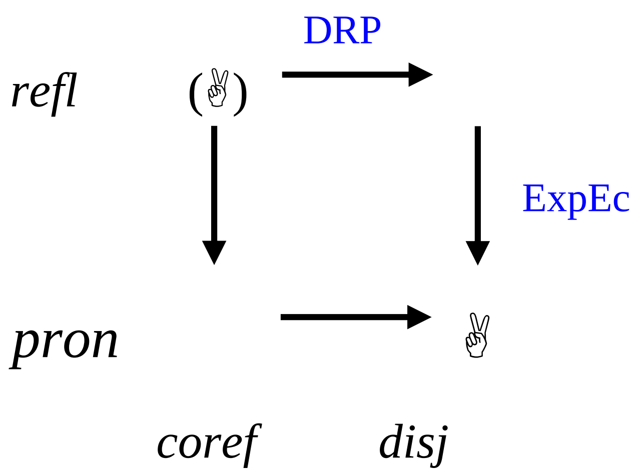
Problems

- Failure of complementary distribution
- Long-distance reflexives
- The existence of languages that lack reflexives. The pronoun allow for both interpretations. (Frisian, Old English, Biblical Hebrew, majority of Australian languages)

Levinson’s B-first Account

DRP (disjoint reference presumption): The arguments of a predicate are disjoint (~~unless marked otherwise~~). Formally: *disj* ←

EXPEC (expressive economy): pronouns are preferred to reflexives (the more marked, prolix forms) Formally: *pron* ←



Problem

For languages *without* reflexives, the DRP can be overridden in certain contexts. (DRP as a **pragmatic** presumption), and the pronoun allows for both interpretations. Why is this not possible for languages *with* reflexives?

Grammaticalization

Levinson: A-first systems evolve from B-first systems

Old-English (B-first) (Visser 1963)

---

(1) No encoded reflexives. Plain Pronouns used reflexively.

(2) *Self* as an emphatic adjectival adjunct … Evolution of *self* as a reflexive marker

(3) The loss of the reflexive use of the ordinary pronoun

Modern English (A-first)

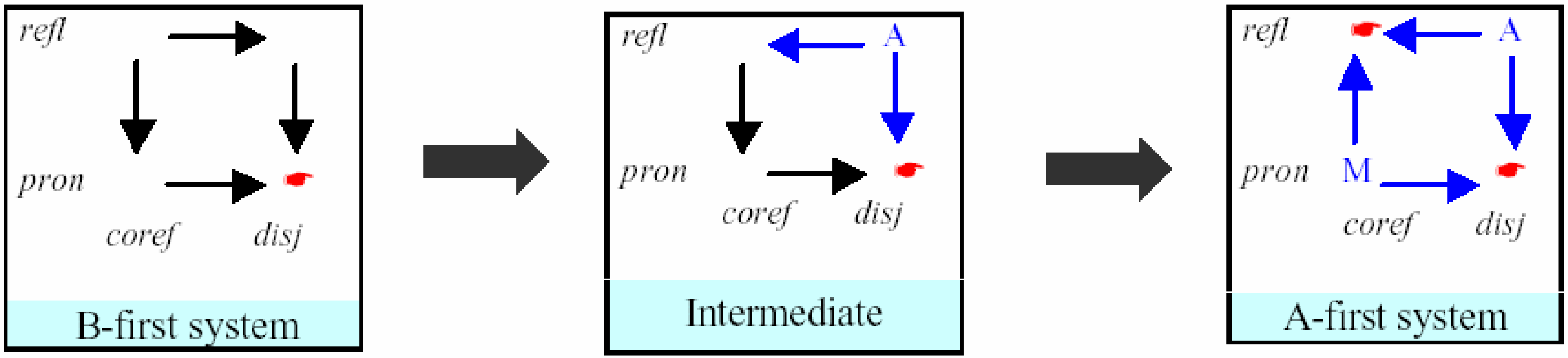
Grammaticalization cont.

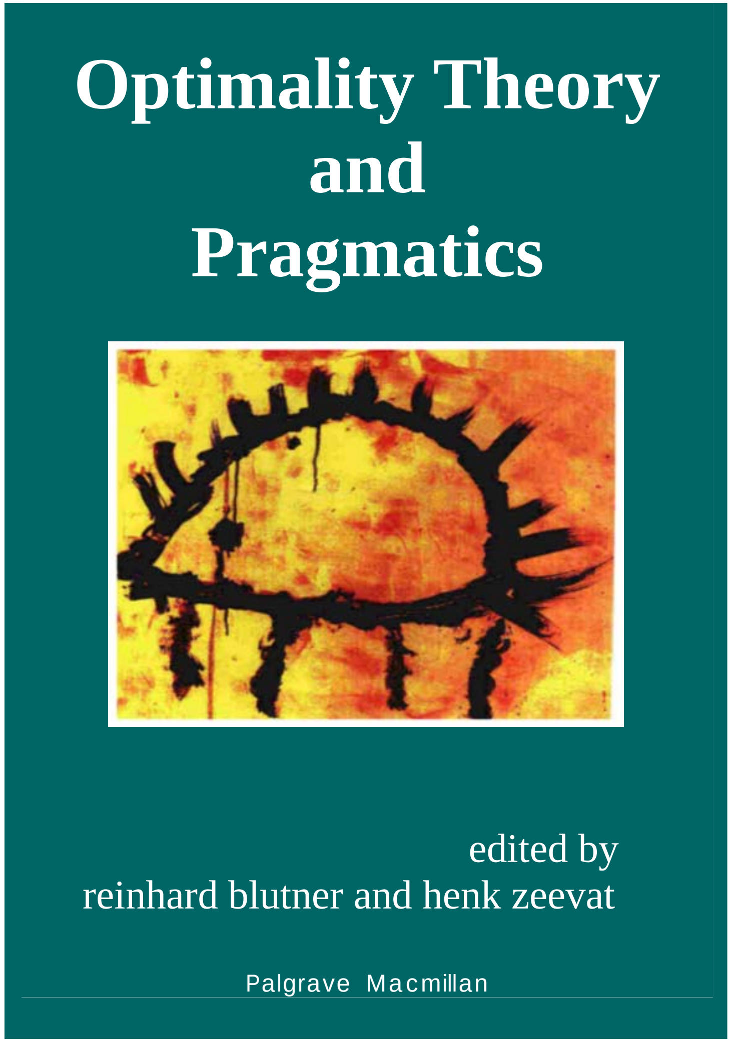
'Grammaticalization' = the harnessing of pragmatics by a grammar (Haiman 1985)

Selection of context-independent contextual constraints – simplified:

PRINC A : *coref* ← *refl*

REF EC (M) : *refl* ← *coref*




Optimality Theory
and
Pragmatics
edited by
reinhard blutner and henk zeevat
Palgrave Macmillan