Variations in Optimality Theory:
Simulated Annealing and other methods

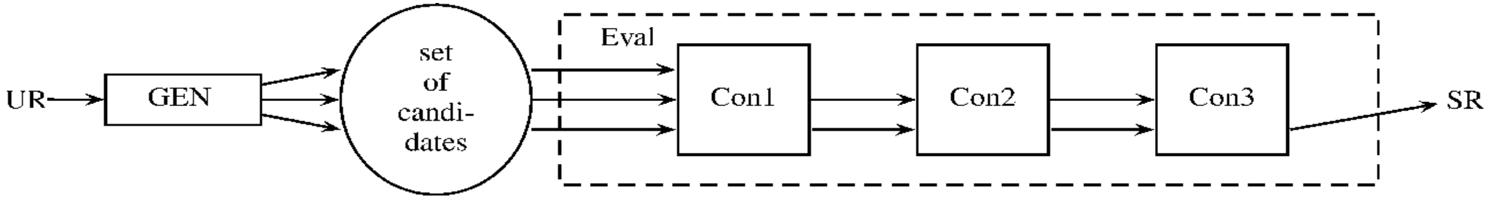
Tamás Bíró, Dicky Gilbers, Maartje Schreuder

Center for Language and Cognition Groningen
University of Groningen

TABU Day

June 3, 2005

Optimality Theory: a model of language typology

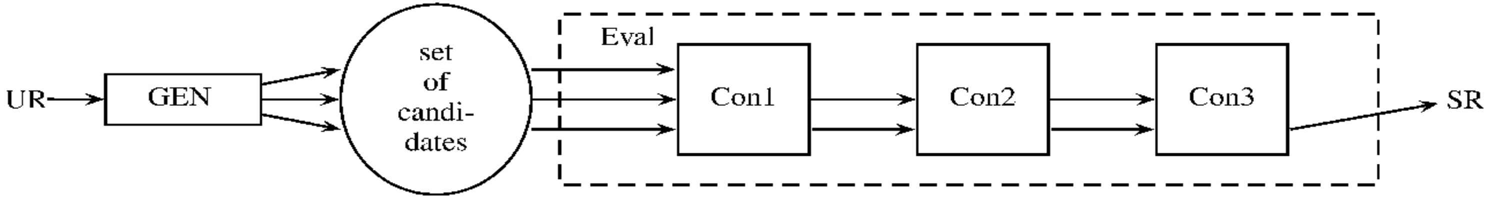


- Candidates = types of language
- Constraint hierarchy
- Thus: one output per hierarchy
- [mɛlk] ∼ [mɛlək]: how to account for variations?

Overview

- Different extensions of OT: [Boersma and Hayes(2001)], [Anttila and Cho(1998)], [Coetzee(2004)]
- An alternative proposal:
  Simulated Annealing for Optimality Theory [Bíró, forthcoming]
- The adequateness of Simulated Annealing for Optimality Theory
- Conclusion

More outputs in Optimality Theory?



For *one* hierarchy, exactly *one* violation profile can be *optimal*.

- One violation profile = more candidates
- One grammar = more hierarchies
- Output = not only the optimal one

Option 1: one grammar = more hierarchies (1)

1/1: Ad hoc (?) reranking of the constraints,

E.g. [Schreuder and Gilbers(2004)]:

Slow (andante) speech:

fototoestel	OOC	*ΣΣ	PRS-σ
☞(fóto)(tòestel)		*	
(fóto)toe(stèl)	*!		*

Fast (allegro) speech:

fototoestel	*ΣΣ	OOC	PRS-σ
(fóto)(tòestel)	*!		
☞(fóto)toe(stèl)		*	*

- Q: Categorical switch to a totally new grammar?

Option 1: one grammar = more hierarchies (2)

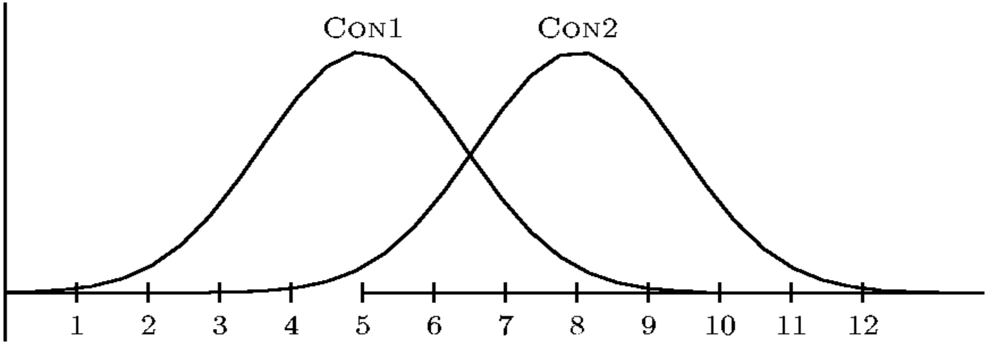
1/2: Unordered constraints [Anttila and Cho(1998)]
All permutations yield their output with equal probability

fototoestel	OOC	*ΣΣ	PRS-σ
☞ (fóto)(tòestel)		*	
(fóto)toe(stèl)	*!		*

- Simple.
- Really free alternations (no control)
- Only certain probability distributions are possible: 50% each.

Option 1: one grammar = more hierarchies (3)

1/3: Stochastic reranking
[Boersma and Hayes(2001)]
Robust learning algorithm
(GLA)



- Most probability distributions possible:
- Two constraints: alternative forms $< 50\%$.
- Same probability of reranking, independently of input.

Option 2: Candidates with the same violation marks

	C1	C2	C3	...	Ci
☞ *cand1*		*	*	...	*
☞ *cand2*		*	*	...	*
cand3		**!		...	**
cand4	*!		*	...	

- There is always a constraint differentiating between them:
- even if low ranked, the analysis does not work.
- Very low ranked constraints are inactive?

Option 3: Non-optimal also as outputs (1)

3/1: critical constraint [Coetzee(2004)]

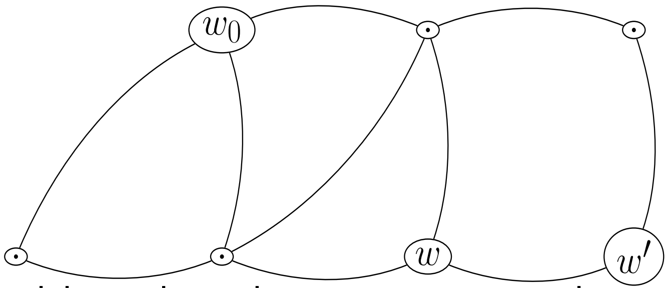
	C1	C2	C3	...	**C-crit**	Ci
☞ *cand1*		*	*	...	*	
$\sim$ *cand2*		*	*	...	*	*
cand3		**!		...	**	
cand4	*!		*	...		*

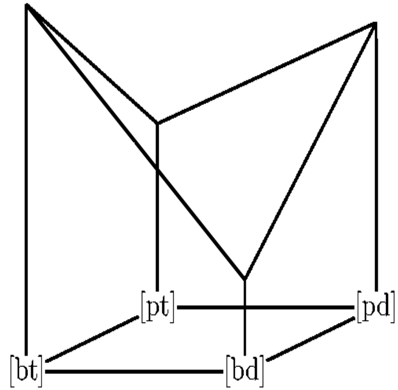
- All candidates surviving the critical constraint appear as alternative forms.

- Non-optimal candidates also ranked: better –> more frequent.

Option 3: Non-optimal also as outputs (2)

3/2: simulated annealing [Bíró, forthcoming]

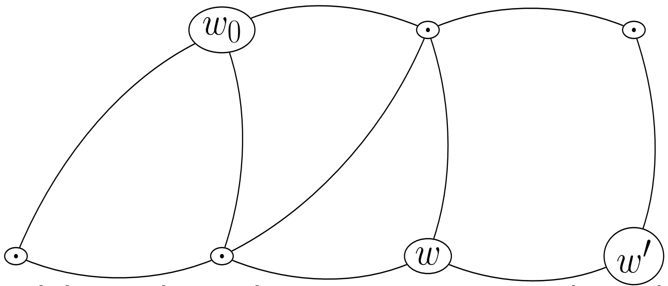


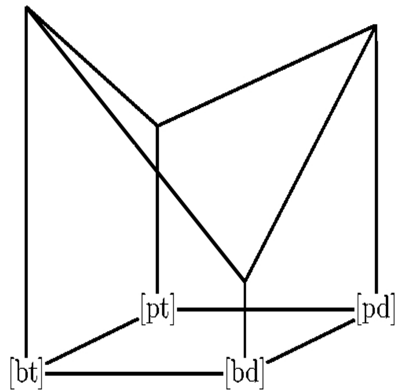


- Neighbourhood structure on the candidate set
- Vertical dimension = harmony; random walk in this landscape
- If neighbour more optimal: move.
- If less optimal: move in the beginning, don't move later

Option 3: Non-optimal also as outputs (3)

3/2: simulated annealing [Bíró, forthcoming]





- Neighbourhood structure –> local optima

- System can get stuck in local optima: alternation forms

- Precision of the algorithm depends on its speed.

- Alternation forms with $> 50\%$ probability can be accounted for.

The Art of Using Simulated Annealing Optimality Theory

- Take a traditional OT model
- Add some *convincing* neighbourhood structure to the candidate set
- Local (non-global) optima = alternation forms
- Run simulation (some more technical details needed...):
  - Slowly: very likely to be returned only the grammatical form
  - Quickly: quite likely to be returned local (non-global) optima

Is Simulated Annealing Better?

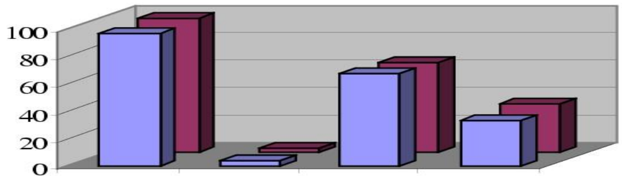
	fototoestel	OOC	*ΣΣ	PRS-σ
☞	(fóto)(tòestel)		*	
$\sim$	(fóto)toe(stèl)	*!		*

Problems with different approaches:

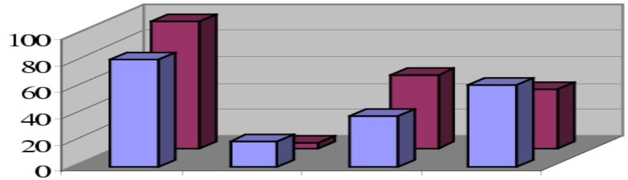
- Categorical reranking: both forms present at higher speech rates?
- Unranked constraints (Anttila): speech rate differences?
- Stochastic OT (Boersma): $> 50\%$ fast speech forms?
- Coetzee: OOC is the critical constraint, thus all forms are $\sim$!
- All approaches: different input – same probabability of reranking

Simulated Annealing Is Better!

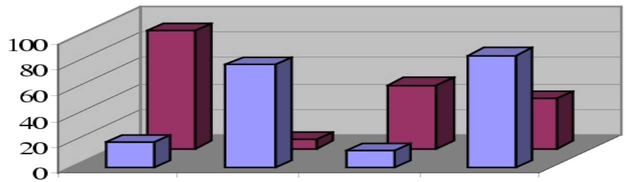
ùit.gè.ve.ríj vs. *ùit.ge.ve.ríj*



stú.die.tòe.la.ge vs. *stú.die.toe.là.ge*



per.fèc.tio.níst vs. *pèr.fec.tio.níst*



[Schreuder (forthcoming) vs. Bíró (forthcoming)]

First two: slow (andante) speech vs. last two: fast (allegro) speech

Simulated Annealing Is Better!

fo.to.toe.stel 'camera'	*uit.ge.ve.rij* 'publisher'	*stu.die.toe.la.ge* 'study grant'	*per.fec.tio.nist* 'perfectionist'
OOC to: susu	ssus	susuu	usus
fó.to.tòe.stel slow: *1.00* fast: *0.82*	*ùit.gè.ve.ríj* slow: *0.97* / **0.96** fast: *0.65* / **0.67**	*stú.die.tòe.la.ge* slow: *0.96* / **0.81** fast: *0.55* / **0.38**	*per.fèc.tio.níst* slow: *0.91* / **0.20** fast: *0.49* / **0.13**
fó.to.toe.stèl slow: *0.00* fast: *0.18*	*ùit.ge.ve.ríj* slow: *0.03* / **0.04** fast: *0.35* / **0.33**	*stú.die.toe.là.ge* slow: *0.04* / **0.19** fast: *0.45* / **0.62**	*pèr.fec.tio.níst* slow: *0.07* / **0.80** fast: *0.39* / **0.87**

Simulated and **observed** frequencies.

[*Bíró (forthcoming)*, **Schreuder (forthcoming)**]

Summary

- Different approaches to variation in OT
- Simulated Annealing is a promising alternative
- OT = competence model: what is an optimal candidate
- SA = performance model: finds optimal candidate
  - A near-good solution returned in constant time
  - You can speed up the algorithm by decreasing precision (cf. fast speech phenomena)

Thank you for your attention!

References

[Anttila and Cho(1998)] Arto Anttila and Young-mee Yu Cho. Variation and change in optimality theory. *Lingua*, 104(1-2):31–56, 1998.

[Boersma and Hayes(2001)] Paul Boersma and Bruce Hayes. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry*, 32:45–86, 2001.

[Coetzee(2004)] Andries W. Coetzee. *What it Means to be a Loser: Non-optimal Candidates in Optimality Theory*. Ph.D. Dissertation, UMass, Amherst, Massachusetts; ROA-687, 2004.

[Schreuder and Gilbers(2004)] Maartje Schreuder and Dicky Gilbers. The influence of speech rate on rhythm patterns. In Dicky Gilbers, Maartje

Schreuder, and Nienke Knevel, editors, *On the Bounderies of Phonology and Phonetics*, pages 183–201. University of Groningen, Groningen, 2004.