Elephants and Optimality Again

SA-OT accounts for pronoun resolution in child language

Tamás Biró

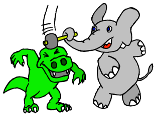
Alfa-Informatica, RUG

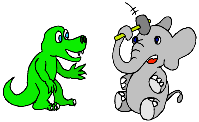
CLIN 19, January 22, 2009

Elephants and Optimality?

Possible correct interpretations of the title:

- Use of tools in explaining cognitive phenomena: should be “optimal”, and not “too heavy”, hitting too strong.
- Optimality Theory: hit the worst candidate.
- Elephants and alligators of pronoun resolution:





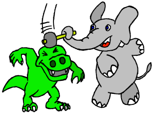
(drawings by Robbert Prins)

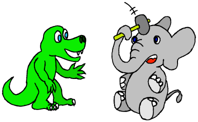
Source: Petra Hendriks, `http://www.let.rug.nl/hendriks/vici.htm`.

Elephants and Optimality?

Possible correct interpretations of the title:

- Use of tools in explaining cognitive phenomena: should be "optimal", and not "too heavy", hitting too strong.
- Optimality Theory: hit the worst candidate.
- Elephants and alligators of pronoun resolution:





(drawings by Robbert Prins)

Source: Petra Hendriks, `http://www.let.rug.nl/hendriks/vici.htm`.

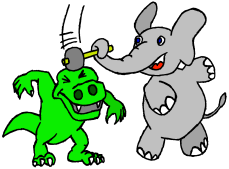
Overview

1. Language acquisition again
2. Simulated Annealing for OT
3. A model for pronoun resolution
4. Simulation results
5. Summary

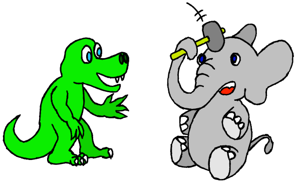
Language acquisition: What do children miss?

- P&P: parameter setting / constraint ranking?
- Principles / constraints?
- A (major) component of the “language device”?
- Performance: working memory, computing power?

Pronoun resolution problem: data



The elephant is hitting him.



The elephant is hitting himself.

Source: Petra Hendriks, `http://www.let.rug.nl/hendriks/vici.htm`.

- "Here you see an elephant and an alligator. Does the elephant hit him?"
- "What does the elephant do?"
- Children of age 4-6 are better at producing pronouns (and reflexives) than interpreting them. Interpretation performance: 50-80 %.

Pronoun interpretation problem: possible explanations

Government and Binding (GB):

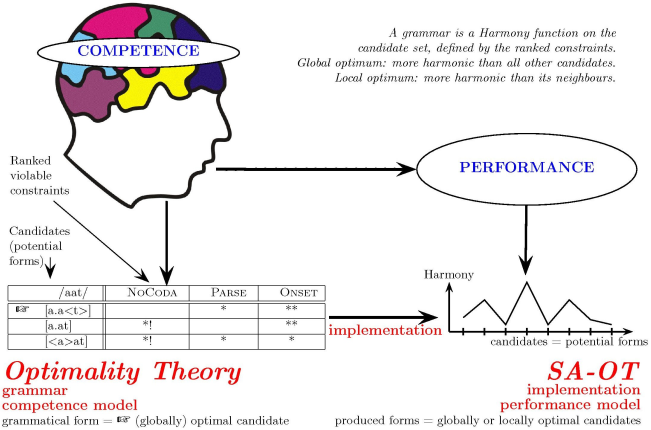
- Principle A: anaphors must be bound within their domain.
- Principle B: pronouns must not be bound within their domain.
- Principle C: R-expressions must not be bound.

- Chien and Wexler: children do not have Principle B yet, due to apparent violations (*He$_i$ looks like him$_j$).*
- Reinhart: insufficient working memory in children to perform necessary computations.
- Hendriks and Spenader: Principle A + bidirectional OT (Principle B not necessary). Children do not have bi-OT before fully developed Theory of Mind.

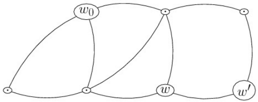
Overview

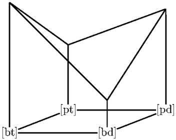
1. Language acquisition again
2. Simulated Annealing for OT
3. A model for pronoun resolution
4. Simulation results
5. Summary

Simulated Annealing for OT (SA-OT)



Simulated Annealing for OT – general idea





- Neighborhood structure on the candidate set.
- Landscape's vertical dimension = harmony; random walk.
- If neighbor more optimal: move.
- If less optimal: move in the beginning, don't move later.
- Neighborhood structure → local optima.
- System stuck in local optima: *performance* errors.
- *Precision* depends on # of iterations.

(Cf. Biro, in: Proc. CLIN 2004.)

Optimality Theory and Harmony Grammar

Objective function to be optimized:

$$H(w) = C_n(w) \cdot q^n + ... + C_i(w) \cdot q^i + ... + C_1(w) \cdot q + C_0(w)$$

- Harmony Grammar: real valued q.
- Optimality Theory: $q = \omega$ or $q \rightarrow +\infty$ (strict domination).

- Harmony Grammar:
  with more iterations, precision converges to 1.
- Optimality Theory: not always!

Overview

A model for pronoun resolution

"Does the elephant hit him?"

- Candidate set 1: *him* refers to
  $\{(\text{alligator}), (\text{elephant}), \emptyset, (\text{alligator\&elephant})\}$.
- Candidate set 2 (with insertion):
  $\{(\text{alligator}), (\text{elephant}), \emptyset, (\text{alligator\&elephant})\} \times \{0, 1, 2, ...\}$.
- Neighborhood structure:
  add/remove one object from the reference set.
- Constraints:
  - PROKNOWN: Reference set must include object from context.
  - AGRNUMBER: reference set cardinality $= 1$.
  - NO3RD: # of inserted elements.
  - PRINCIPLEB: *elephant* not in reference set.
- Hierarchy: PRO $\gg$ AGRNUMBER $\gg$ NO3RD $\gg$ PRINCIPLEB.

A model for pronoun resolution

	PRO	AGRNUMBER	NO3RD	PRINCIPLEB
0	1	1	0	0
∼ e	0	0	0	1
ea	0	1	0	1
☞ a	0	0	0	0
0 +1	1	0	1	0
e +1	0	1	1	1
ea +1	0	1	1	1
a +1	0	1	1	0
...				
0 +k	1	1	k	0
e +k	0	1	k	1
ea +k	0	1	k	1
a +k	0	1	k	0
...				

From Harmony Grammar to Optimality Theory

Candidate set 1 (no insertion), $K_{max} = 5$, $T_{step} = 0.1$.

q	precision
OT	0.500
30	0.499 ± 0.008
20	0.500 ± 0.012
10	0.499 ± 0.003
5	0.511 ± 0.001
3	0.550 ± 0.005
2.5	0.580 ± 0.003
2.0	0.633 ± 0.003
1.8	0.666 ± 0.003
1.7	0.687 ± 0.007
1.6	0.716 ± 0.006
1.5	0.749 ± 0.008

q	precision
1.4	0.790 ± 0.004
1.3	0.847 ± 0.001
1.2	0.911 ± 0.002
1.15	0.945 ± 0.003
1.10	0.978 ± 0.001
1.08	0.986 ± 0.001
1.06	0.994 ± 0.001
1.05	0.997 ± 0.001
1.04	0.9985 ± 0.0003
1.03	0.9991 ± 0.0005
1.02	0.99977 ± 0.00015
1.01	0.99997 ± 0.00006

Tuning the parameters in Optimality Theory

Candidate set 2 (with insertion), $T_{step} = 0.1$.

K_{max}	precision
1	0.575 ± 0.003
3	0.616 ± 0.004
5	0.649 ± 0.003
8	0.684 ± 0.002
10	0.700 ± 0.007
30	0.798 ± 0.003
50	0.839 ± 0.003
80	0.871 ± 0.004
100	0.881 ± 0.003
300	0.929 ± 0.003
500	0.945 ± 0.002
800	0.954 ± 0.002
1000	0.961 ± 0.005
2000	0.972 ± 0.002

(NB: More explanation in Bíró (2006).)

From Harmony Grammar to Optimality Theory

Candidate set 2 (with insertion), $K_{max} = 5$, $T_{step} = 0.1$.

q	precision
OT	0.649 ± 0.003
30	0.659 ± 0.003
20	0.664 ± 0.008
10	0.647 ± 0.002
5	0.641 ± 0.002
3	0.634 ± 0.002
2.5	0.632 ± 0.004
2.0	0.648 ± 0.003
1.8	0.671 ± 0.006
1.7	0.680 ± 0.006
1.6	0.704 ± 0.001
1.5	0.725 ± 0.005

q	precision
1.40	0.761 ± 0.006
1.30	0.804 ± 0.005
1.20	0.872 ± 0.003
1.15	0.910 ± 0.003
1.10	0.949 ± 0.002
1.08	0.963 ± 0.001
1.06	0.978 ± 0.002
1.05	0.983 ± 0.001
1.04	0.990 ± 0.002
1.03	0.993 ± 0.0004
1.02	0.9967 ± 0.0006
1.01	0.9989 ± 0.0004

Summary

Simulated Annealing with *Optimality Theory/Harmony Grammar* provides a framework to account for delay in the pronoun interpretation problem. Adults make less “performance errors” than children:

- Learning social cognition, etc.: enlarge the candidate set with candidates including not present elements.
  (*Godot-effect*: crucial role played by “invisible” candidates.)
- More computational power: use of higher K_{max}.
- Learn to use a more flexible grammar for semantic-pragmatic issues: no more strict domination, reduce q.

NB: I’m not arguing against previous explanations!
Future work to compare them.

Thank you for your attention!

Tamás Biró
http://www.let.rug.nl/birot, birot@nytud.hu