Finding the Right Words
Implementing Optimality Theory with Simulated Annealing

Tamás Bíró
Humanities Computing, CLCG, University of Groningen, The Netherlands birot@birot.hu

ABSTRACT

Simulated Annealing (SA), a standard heuristic algorithm for solving optimisation problems (Kirkpatrick et al., 1983), is employed to find the optimal candidate in the candidate set, as defined by *Optimality Theory* (OT; Prince & Smolensky, 1993/2004). This optimal candidate is predicted by OT to be the grammatical form. Simulated Annealing does not always find the correct solution, still returns some result within a constant time. Erroneous outputs of SA correspond to performance errors. Accelerating the algorithm is possible by giving up on precision: a faster algorithm produces more speech errors. Thus, the algorithm developed, *Simulated Annealing for Optimality Theory* (SA-OT) models linguistic performance, built upon Optimality Theory as a competence model.

Competence and performance: a novel view

1. *Competence*: the static knowledge	grammatical	(explained by) grammar
2. *Mental computation* in the brain	produced	implementation of grammar
3. *Performance* in its outmost sense	produced	phonetics, pragmatics, etc.

Proposal: Performance errors – forms produced but not grammatical – are partially due to errors in the mental computation. It is the *implementation of the grammar* that has to account for them.

Similar proposal, developed independently but left untested, in Smolensky et al.'s *Harmonic Mind* (2006, vol. 1, p. 229).

Case 1: fast speech

- *Grammatical form*: whose frequency decreases in fast speech.
- *Fast speech form*: whose frequency increases in fast speech.

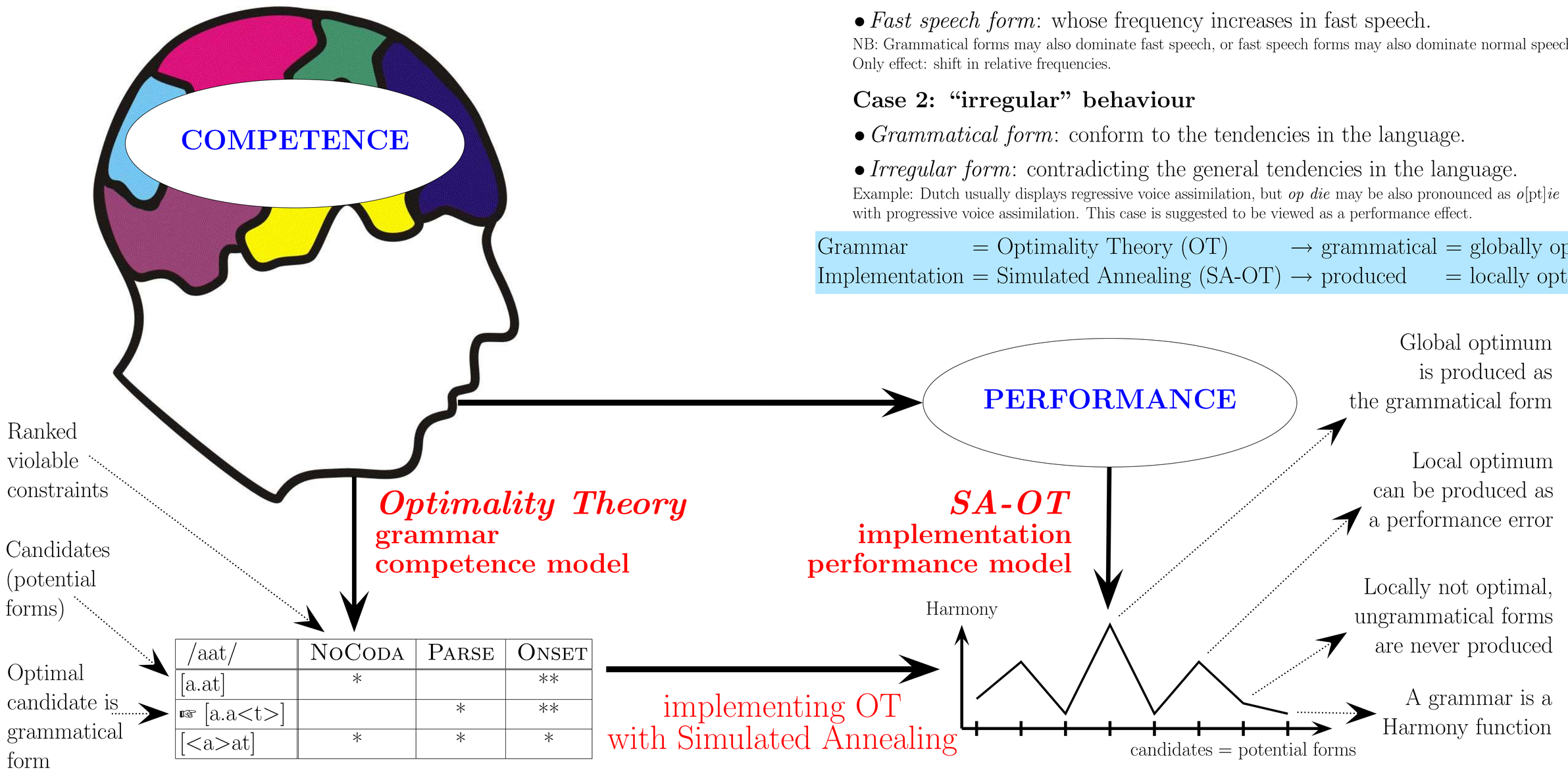
NB: Grammatical forms may also dominate fast speech, or fast speech forms may also dominate normal speech. Only effect: shift in relative frequencies.

Case 2: "irregular" behaviour

- *Grammatical form*: conform to the tendencies in the language.
- *Irregular form*: contradicting the general tendencies in the language.

Example: Dutch usually displays regressive voice assimilation, but *op die* may be also pronounced as *o*[pt]*ie* with progressive voice assimilation. This case is suggested to be viewed as a performance effect.

Grammar	= Optimality Theory (OT)	→ grammatical	= globally optimal
Implementation	= Simulated Annealing (SA-OT)	→ produced	= locally optimal



Simulated Annealing

Originating in physics, Simulated Annealing (Boltzmann Machines or stochastic gradient ascent; Kirkpatrick et al., 1983), is a widespread heuristic technique for combinatorial optimisation. A random walk is performed on the search space until it is trapped in the global or in another local optimum. (A local optimum is a point that is better than all its neighbours). The slower the speed of the algorithm, the higher the chance of finding the global optimum.

Variation in Optimality Theory

Variation in languages can be either free, or dependent upon factors like register or speech rate. A traditional competence model simply predicts both forms to be grammatical. Yet, phenomena such as fast speech consist of shifts in the relative frequencies. Thus, linguistic models should also account for these frequency changes. OT offers several ways to map one underlying form to more surface forms:

1. Alternating forms assigned the same violations.
2. Several hierarchies in one grammar: reranking, Anttila, Boersma, Reynolds.
3. Non-optimal forms also emerging: Coetzee, Simulated Annealing (SA-OT).

SA-OT gives better frequency predictions than most other approaches for fast speech. Moreover, by distinguishing between competence and performance, SA-OT needs not postulate a slightly different OT grammar, a different competence, for each speech rate. SA-OT makes directly interpretable predictions by leaving the competence model intact and by altering only the parameters of the performance model. Fast speech phenomena emerge from running the implementation faster.

Example: Dutch metrical stress in fast speech

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

Simulated (SA-OT) / **observed** (M. Schreuder and D. Gilbers, 2004) frequencies.
In SA-OT: $t_{step} = 3$ used for fast speech and $t_{step} = 0.1$ for slow speech.

The broader cognitive context

SA-OT explains why certain speech errors are made and why not others. In doing so, it belongs to the *heuristic* mental computation models, describing a wide range of cognitive tasks. Supposedly, the human mind is willing to make some errors in order to achieve a fast but reasonably good performance. That is why *errare humanum est* – To Err Is Human. And not only in finding the right words.

References

S. Kirkpatrick et al (1983): *Optimization by Simulated Annealing*, Science 220.
A. Prince and P. Smolensky (1993): *Optimality Theory: Constraint Interaction in Generative Grammar*, RuCCS-TR2.
M. Schreuder and D. Gilbers (2004): *The Influence of Speech Rate on Rhythm Patterns*, in: *On the Bounderies of Phonology and Phonetics*, Groningen.