# 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 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).

### Example: Fast speech: Dutch metrical stress

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

**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 dominate also 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 pronounced as *o*[pt]*ie* with progressive voice assimilation: suggested to view as performance effect.

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


PERFORMANCE

Ranked violable constraints

Candidates (potential forms)

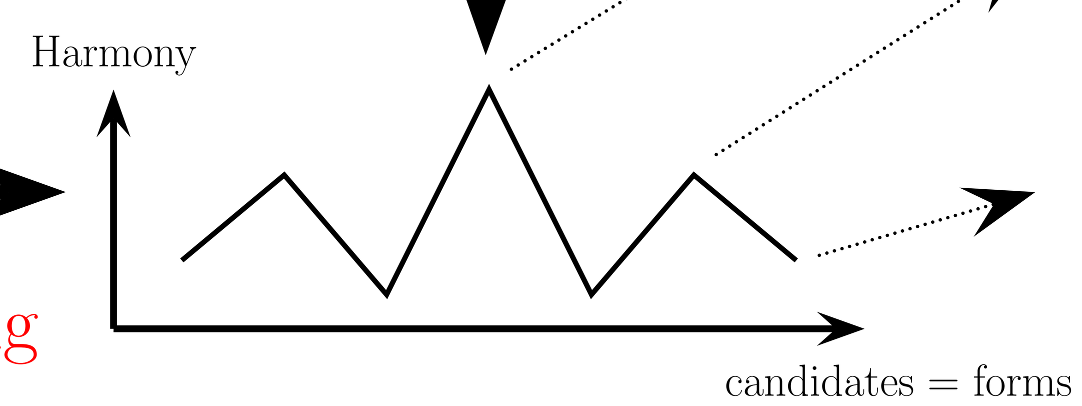
Optimal candidate is grammatical form

**Optimality Theory**
grammar
competence model

**SA-OT**
implementation
a performance model

| /aat/ | NoCoda | Parse | Onset |
|---|---|---|---|
| [a.at] | * | | ** |
| ☞ [a.a<t>] | | * | ** |
| [<a>at] | * | * | * |

implementing OT with Simulated Annealing




Global optimum is produced as the grammatical form

Local optimum can be produced as a performance error

A grammar is a Harmony function

## Variation in Optimality Theory

Variation in languages can be either free, or dependent upon factors such as register or speech rate. A traditional competence model simply predicts both forms to be grammatical. Yet, some phenomena, such as fast speech, only consist of shifts in the relative frequencies. Therefore, linguistic models should also account for these frequency changes.

Optimality Theory offers several ways to map one underlying form to more surface forms:

1. Alternating forms assigned the same violations ("the poor man's way of dealing with variation").
2. Several hierarchies within one grammar: e.g., ad hoc reranking, Anttila, Boersma, Reynolds.
3. Non-optimal forms also emerging: Coetzee, Simulated Annealing (SA-OT).

We argue that 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.

## Representations of the Harmony function

The *Harmony function* $H(w)$ measures the "goodness" of candidate $w$ in OT. Derived from the number of violation marks assigned by each constraint to $w$, $H(w)$ can be represented as:

| | NoCoda | Parse | Onset |
|---|---|---|---|
| $w$ | * | | ** |

1. A multiset (in which multiplicity is significant) of violations: $H(w) = \{\textsc{NoCoda}, \textsc{Onset}, \textsc{Onset}\}$.
2. A vector (a row in an OT tableau): $H(w) = (\textsc{NoCoda}(w), \textsc{Parse}(w), \textsc{Onset}(w)) = (1, 0, 2)$.
3. A real number, using a set of weights $s_i$ (higher ranked constraints have higher weights): $H(w) = \textsc{NoCoda}(w) \cdot s_{\textsc{NoCoda}} + \textsc{Parse}(w) \cdot s_{\textsc{Parse}} + \textsc{Onset}(w) \cdot s_{\textsc{Onset}} = 1 \cdot 16 + 0 \cdot 4 + 2 \cdot 1 = 18$ (with weights 16, 4 and 1) → only if the number of violation marks has an upper bound!
4. A polynomial (a function of $q$, and $q$ goes to infinity, to account for *strict domination*): $H(w)[q] = \textsc{NoCoda}(w) \cdot q^3 + \textsc{Parse}(w) \cdot a^2 + \textsc{Onset}(w) \cdot q = 1 \cdot q^3 + 0 \cdot q^2 + 2 \cdot q$.
5. An ordinal number (a mathematically exact way of counting with infinite values): $H(w) = \omega^3 \cdot \textsc{NoCoda}(w) + \omega^2 \cdot \textsc{Parse}(w) + \omega \cdot \textsc{Onset}(w) = \omega^3 \cdot 1 + \omega^2 \cdot 0 + \omega \cdot 2$.

Simulated Annealing is adapted to Optimality Theory using more of these representations, and in each case we obtain the same algorithm: *Simulated Annealing for Optimality Theory* (SA-OT).

## 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 being trapped in the global or in another local optimum. The slower the speed of the algorithm, the higher the chance of finding the global optimum.

### Simulated Annealing for Optimality Theory (SA-OT) Algorithm

```
ALGORITHM: Simulated Annealing for Optimality Theory
Paramters: w_init, K_max, K_min, K_step, t_max, t_min, t_step
        # t_step: number of iterations / speed of simulation
w <-- w_init ;
   for K = K_max to K_min step K_step
        for t = t_max to t_min step t_step
             choose random w' in neighbourhood(w) ;
             calculate < C , d >  = ||H(w')-H(w)|| ;
                if d <= 0 then w <-- w'
             else           w <-- w' with probability
                         P(C,d;K,t) = 1          , if C < K
                                    = exp(-d/t)  , if C = K
                                    = 0          , if C > K
        end-for
   end-for
return w
```

## 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-TR-2.

M. Schreuder and D. Gilbers (2004): *The Influence of Speech Rate on Rhythm Patterns*, in: *On the Bounderies of Phonology and Phonetics*, Groningen.