

Summary

This dissertation presents an implementation of *Optimality Theory* (OT, Prince and Smolensky, 1993) that also aims at accounting for certain variations in speech. The *Simulated Annealing for Optimality Theory Algorithm* (SA-OT, Fig. 2.8, on page 64) combines OT with *simulated annealing*, a widespread heuristic optimisation technique. After a general introduction to Optimality Theory and the discussion of certain “philosophical background questions” (especially on the role of probabilities in linguistics; Chapter 1), the SA-OT Algorithm is introduced (informally in section 2.2, mathematically in sections 3.3 and 3.4), put into a broader context (section 2.1, Chapter 4, and sections 8.2 and 8.3), and experimented with (section 2.3, Chapters 5-7).

Reeves (1995) defines *heuristic* as “a technique which seeks good (i.e. near-optimal) solutions at a reasonable computational cost without being able to guarantee either feasibility or optimality, or even in many cases to state how close to optimality a particular feasible solution is.” Even if they are not exact, these algorithms are very useful in solving efficiently hard computational problems, similar to the task of finding the optimal candidate in an OT candidate set. A *good* solution suffices in many applications, and there is no need to allocate huge computational resources to find the *best* solution. As section 2.1 argues, heuristic algorithms—such as SA-OT—may serve as adequate models of the computations performed by the human brain for at least three reasons: (1) many of these algorithms are simple, (relatively) efficient and produce *some* output within a predefined time span, even if (2) they may make errors, and finally (3) the algorithm can be speeded up with a price to be paid in reduced precision. A faster computation is possible, but more prone to make errors. The adequacy of such a model is corroborated if besides the grammatical forms it also reproduces the empirically observable error patterns under different conditions. Importantly, these predictions are quantitative, and the algorithm’s parameters can “fine-tune” the output frequencies of the erroneous or alternating forms.

Table 2.1 (page 43) formulates this idea: by distinguishing between a linguistic model and its implementation, one can account for both linguistic competence and certain types of linguistically motivated performance phenomena. Thus an adequate linguistic model (a *grammar*, such as a well-founded OT grammar) predicts correctly which forms are judged as grammatical by the native speaker. This layer refers to the *static knowledge of the language* in the native speaker’s brain. On top of that is built the *implementation of the grammar* as a model of the *dynamic language production process*. Similarly to human speech, the implementation of the grammar need not be exact, but the errors made by the implementation should correspond to the observed performance errors.

In particular, SA-OT requires a *topology* (a *neighbourhood structure*) on the OT candidate set. Consequently, the notion of a *local optimum* is introduced: a candidate that is more harmonic than all its neighbours is a local optimum, independently of whether it is the most harmonic element of the entire candidate set. Local optima are the candidates that can emerge as outputs in SA-OT. The *global optimum* predicts the grammatical form, whereas all other outputs should model performance errors.

How does the SA-OT Algorithm work? A random walk is performed on the candidate set. In each iteration, the random walker chooses randomly a candidate w' among the neighbours of its present position w . For instance, a minimal *basic operation* transforms w into w' . The definition of the topology on the candidate set also includes the *a priori probabilities* $P_{choice}(w'|w)$, which determine the chance of choosing w' with the condition that the random walker is in w . Then, the random walker truly moves to w' with some other *transition probability* $P(w \rightarrow w'|T)$ —which depends on the harmonies (violation profiles) of candidates w and w' —else it stays in w .

The *a priori probabilities* $P_{choice}(w'|w)$ do not depend on the violation profiles of the candidates (hence, on the constraint ranking) and are constant during the simulation. The *transition probabilities* $P(w \rightarrow w'|T)$, however, diminish as a function of the parameter T (called *temperature*), gradually, from 1 at the beginning of the simulation to 0 at its end—if w' is less harmonic than w . Otherwise, the *transition probability* from w to w' remains 1, so the random walker is always allowed to move to a better neighbour. (See equation (2.2) on page 39 for this idea in traditional simulated annealing, and equation (2.17) on page 63 and the subsequent “Rules of moving” for SA-OT.) Thus, the random walker will be stuck in some local optimum by the end. The output of the algorithm is the final position of the random walker, and its precision is the likelihood of this local optimum being also globally optimal. Frequently—but certainly not always, as Chapter 6 demonstrates—a slower pace of diminishing the parameter T (that is, a larger number of iterations performed, a slower *cooling schedule*) results in a reduced chance of being stuck in a local optimum that is not globally optimal.

Combining simulated annealing with Optimality Theory has been far from trivial. Traditional simulated annealing optimises a real-valued target function, which is different from the Harmony function employed in OT. In order to introduce the *transition probabilities*, first the difference of two violation profiles has to be defined, and then temperature T is introduced as a pair $\langle K, t \rangle$. The idea is first presented in section 2.2 in a relatively informal way, whereas Chapter 3 argues for the same algorithm by making use of several mathematical formalisms. Chapter 3 starts with the mathematical definition of OT—also in order to show which assumptions are needed in OT and what assumptions can be generalised in future research—followed by a discussion on how to realise the Harmony function using polynomials, on the one hand, and using ordinal numbers, on the other. Both approaches lead to the same way of combining OT with simulated annealing.

The following chapter speculates about two “hot topics” in linguistics: the lexicon and learnability. Apart from trains of thought that are left open for future research and do not play a central role in this dissertation, it introduces a new and formal definition of *Output-Output Correspondence*, or rather *Constituent-Output Correspondence*, which is used in the subsequent chapter.

The remaining three chapters before the conclusion present how concrete linguistic phenomena could be tackled within the framework of SA-OT. In these chapters the emphasis lies not so much on the phonological details of the specific analyses, rather on the methodological problems raised by SA-OT. The experiments with the models demonstrate the role of certain parameters of the algorithm, as well as the importance of certain decisions that have to be made when one decides to use SA-OT. Additionally, different techniques and tricks, different ways of experimenting with the models and different ways of understanding their behaviour are presented throughout these chapters. Section 8.1 summarises in details the various methodological issues dealt with in these chapters.

Using a few observations on Dutch metrical stress assignment, Chapter 5 demonstrates how fast speech phenomena can be reproduced by varying the speed of the algorithm. If the algorithm is run more quickly (with fewer iterations, with a faster *cooling schedule*), then the frequencies of the alternating forms change similarly to the way they change when moving from slow (normal, careful) speech to fast speech, as reported by laboratory experiments.

The toy models in section 2.3 advanced some of the problems that are further dealt with in Chapter 6, using the example of Dutch regressive and progressive voice assimilation. Two models are presented, the first one involving a very simple and restricted candidate set, and the second one displaying an infinite topology. The landscapes, that is, the topologies of the OT candidate set with the Harmony function, were simpler in these models than in the previous chapter, and therefore an analytical discussion of the behaviour of the models could also be included, besides the experimentation with the parameters. Furthermore, these models demonstrated that in the case of SA-OT—unlike in the case of traditional simulated annealing—increasing the number of iterations (having a slower cooling schedule) does not always necessarily lead to an increased probability of returning the globally optimal candidate. Supposing that SA-OT is indeed an adequate model for speech production, this observation opened the floor to speculations about how to keep a grammar simple while still accounting for “irregularities”. Namely, the “irregular” forms are conjectured to be the erroneous outputs (the non-global local optima) that the dynamic language production process cannot avoid producing under any condition.

Chapter 7 discusses two phenomena, both related to word or syllable structure. First, the cliticisation of the article in Hungarian is accounted for, as a function of the speech rate and of the allomorph chosen. The topology of the model has an overall structure that makes it similar to the second topology analysed in Chapter 6. The same type of topology is enriched in the last model presented in this dissertation, which is a first and preliminary attempt to implement the classical OT paradigm for syllabification (basic CV Theory) using SA-OT. So the simulations presented in Chapters 6 and 7 also demonstrate how a certain type of model—which, I conjecture, might become important in future research—can be made gradually more complex. In the same chapters we see how never winning (loser) candidates can still influence the output frequencies and should therefore be included in the candidate set (the *Godot effect*). Moreover, section 7.1 demonstrates how SA-OT can supply arguments for ranking constraints that could not be ranked based on the arguments used in traditional OT.

The concluding Chapter 8 summarises the results of the present thesis and

compares SA-OT to other OT approaches to linguistic variation. Future research should decide whether SA-OT or its competitors, the already existing stochastic OT models are more fruitful, but I believe that they may complement each other. Moreover, future research should also work out certain details, which have been judged so far by many readers as *ad hoc*, in a more persuasive manner. Finally, in section 8.3, SA-OT is put into the context in which Optimality Theory was born more than a decade ago, namely, the cognitive sciences.