How could one do a statistical test of Optimality Theory?*

Greg Kochanski

http://kochanski.org/gpk

January 25, 2005

1 Introduction

Optimality Theory (OT) has been a successful idea in Linguistics. It has spawned hundreds of papers and dozens of Ph.D. dissertations. Unfortunately, despite it's success in that sense, it has never been subjected to any serious test of its validity. In this lecture, we'll look at possibilities for testing OT, why one might want to, and the statistical techniques that would be needed.

2 What is a Theory?

Optimality Theory appears to be science. Complex sets of rules are developed, based on computer programs and data. But is it?

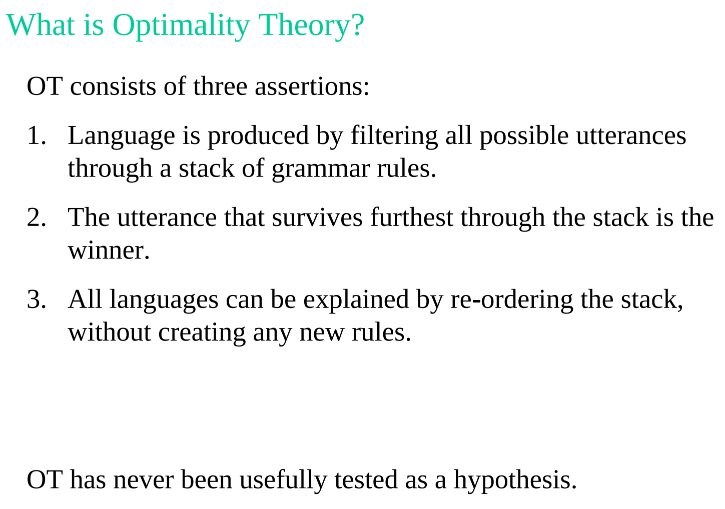
A theory, at least if one is talking about science, provides an explanation of a corner of the natural world. But more than just providing an explanation, a theory should start with premises that are likely to be true and proceed by way of plausible mechanisms. Ideally, it would make predictions about the world that can be tested.

Science proceeds by a mixture of insight and elimination. Insight to find the right answer and elimination to remove all the failed insights that didn't work out. Requiring that theories be produce testable predictions is strictly a pragmatic labour-saving device: many theories that are tested fail, and if that happens they can thereafter be ignored. The scientists who were working on the failed theories can then go off and study something else, and the universities can then spend expensive classroom hours teaching other things, things that are more likely to be true.

More than just making predictions that can be tested, a theory needs to make predictions where the truth is not yet known. Again, from a strictly practical point of view, theories need to have the possibility of failure, otherwise science would get clogged with an accumulated sludge of ideas. Absent failures, there would simply be too much for students to learn.

In Linguistics, the concept of a theory is somewhat broadened to include explanations of the universe that may not be directly testable, but at least have the virtue of being elegant and compact ways of describing part of the world.

The two meanings are actually related. The second meaning includes the explicit and potentially testable assertion that the description is compact, in other words that describing the theory takes fewer words than simply listing all the data. One might use some of the techniques of Information Theory (Shannon etc XXX) to quantify the complexity of the theory and the information present in the



What is Optimality Theory?

OT consists of three assertions:

1. Language is produced by filtering all possible utterances through a stack of grammar rules.
2. The utterance that survives furthest through the stack is the winner.
3. All languages can be explained by re-ordering the stack, without creating any new rules.

OT has never been usefully tested as a hypothesis.

Figure 1: What is Optimality Theory?

data. Ideas such as minimum-descriptor-length codes in computer science attempt to apply this idea to compactly representing data.

A theory might fail this compactness test in a formal sense of requiring too many bits, but it can also fail it in the practical sense that working scientists find that it is simply less bother to just list the data.

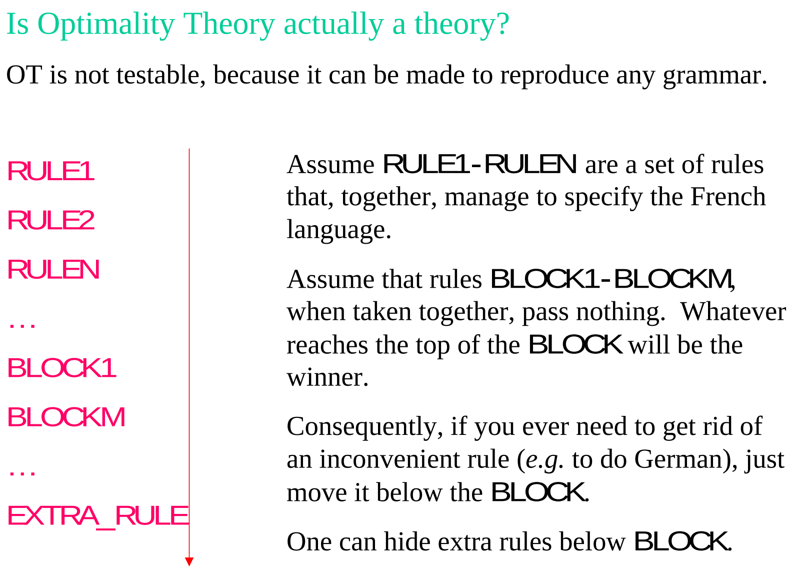
3 Optimality Theory

Optimality theory (REF XXX) is a way to compute phonological representations. In OT, each language is represented as an ordered list of constraints (Figure 1). Each constraint specifies some property of the language: for instance that all syllables are CVC[1]. In OT, one imagines that all possible phonological implementations are poured into the top of this list of constraints. Each constraint acts as a filter, passing only those utterances that meet the constraint on down to the next constraint. OT claims that the legal utterances of a language are those that make it farthest down the stack.

Further, OT claims that all languages share a common set of constraints, so that any language can be represented by re-ordering the list of constraints.

This sounds quite promising as a candidate for a theory. Representing languages as a stack of constraints might well be an elegant and compact way of describing the diversity of language. More importantly, OT makes a prediction that appears strong and testable: any language can be represented by re-ordering the list. Unfortunately, OT is not testable, in the sense that it cannot fail to represent any data. It is too vague and broad to be a theory in that sense.

[1] Many publications on OT involve constraints that are rather more abstract than the requirement of CVC syllables.



Is Optimality Theory actually a theory?

OT is not testable, because it can be made to reproduce any grammar.

RULE1

RULE2

RULEN

...

BLOCK1

BLOCKM

...

EXTRA_RULE

Assume RULE1-RULEN are a set of rules that, together, manage to specify the French language.

Assume that rules BLOCK1-BLOCKM, when taken together, pass nothing. Whatever reaches the top of the BLOCK will be the winner.

Consequently, if you ever need to get rid of an inconvenient rule (*e.g.* to do German), just move it below the BLOCK.

One can hide extra rules below BLOCK.

Figure 2: Hiding Rules in Optimality Theory

3.1 Optimality Theory – Hiding Language-Specific Constraints

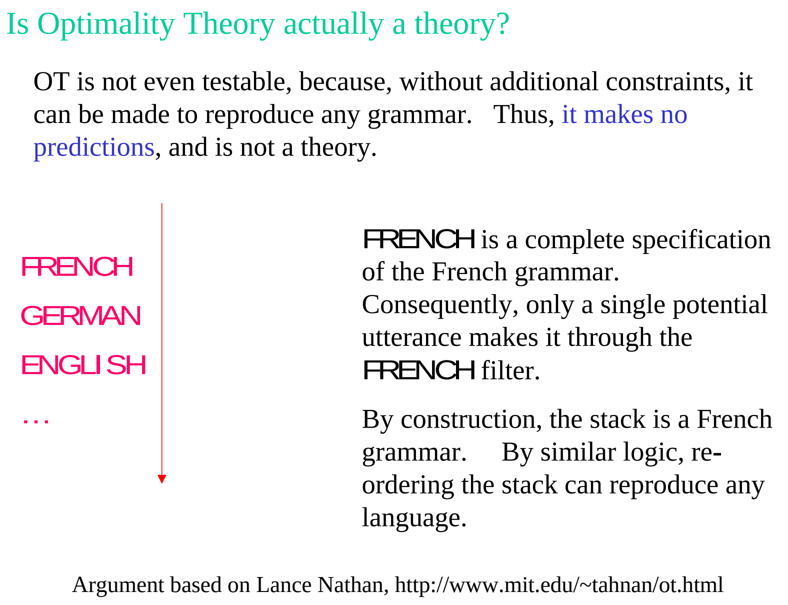
Since the nature of the constraints are not specified in OT, and we are free to invent our own, we can invent constraints like **+FRENCH** to make each language work. Researchers who are too squeamish to compress an entire language into one constraint can split **+FRENCH** into any desired number of fine-grained constraints[2]. Many of these fine-grained components of **+FRENCH** can doubtless be quite generic, appearing in language after language.

However, if generic constraints are not sufficient to explain a certain language, it is possible to introduce language-specific constraints then "hide" them deep down in the OT tableau where they will not cause trouble when one is doing OT on other languages. By this means, a clever theorist can get around OT's claim to represent all languages with the same set of constraints.

For instance, Figure 2 shows an example of hiding a rule below a block in a tableau. In the figure, Rule$_1$ to Rule$_N$ specify the French language. Let us imagine that to specify English, we need a special rule, **EXTRA**. Let us further imagine that the theorist can't figure out where to put **EXTRA**. Imagine that no matter where it is placed among Rule$_1 \ldots$Rule$_N$, it breaks something.

No problem. The theorist needs only to create a block in the tableau, here written as Block$_1 \ldots$Block$_M$. The block is designed so that at most one implementation can pass through. (In a sense, this block is unnecessary, since if the rules for French above the block are properly designed, only one implementation will get to the block anyway.) Below the block, there is certainly at most one implementation.

[2] It is always possible to split a constraint X into two fine-grained constraints: one can write the OT tableau as a series of boolean matrix multiplications where the basis is *reject* followed by a list of all the possible implementations. Each constraint multiplies this vector. The question of splitting a constraint then reduces to the the solvability of the matrix equation $AB = C$, if one is given C. Since matrix multiplication has an identity operator, a solution is always possible; generally, an infinite number of solutions are possible.



Is Optimality Theory actually a theory?

OT is not even testable, because, without additional constraints, it can be made to reproduce any grammar. Thus, it makes no predictions, and is not a theory.

FRENCH
GERMAN
ENGLISH
...

FRENCH is a complete specification of the French grammar. Consequently, only a single potential utterance makes it through the FRENCH filter.

By construction, the stack is a French grammar. By similar logic, re-ordering the stack can reproduce any language.

Argument based on Lance Nathan, http://www.mit.edu/~tahnan/ot.html

Figure 3: OT is not falsifiable.

Since we have constructed a tableau where the implementation is unique below Block_M, nothing below that point can affect the ultimate selection. Consequently, we can hide anything we want below that point. **EXTRA** can be safely placed below.

By symmetry, if we need some language-specific rules to implement French that could cause trouble in implementing English (e.g. **SUPPLÉMENT**), we can likewise hide those below the block.

That means OT's hope to represent all languages *with the same set of constraints* is empty. It says nothing about the unity of human languages because one can always hide the disunities below a block. Consequently, OT is not a theory in the stronger sense: it cannot be proven false. There is no language, whether human, alien, computer, or indeed simply a random list of symbols that cannot be represented.

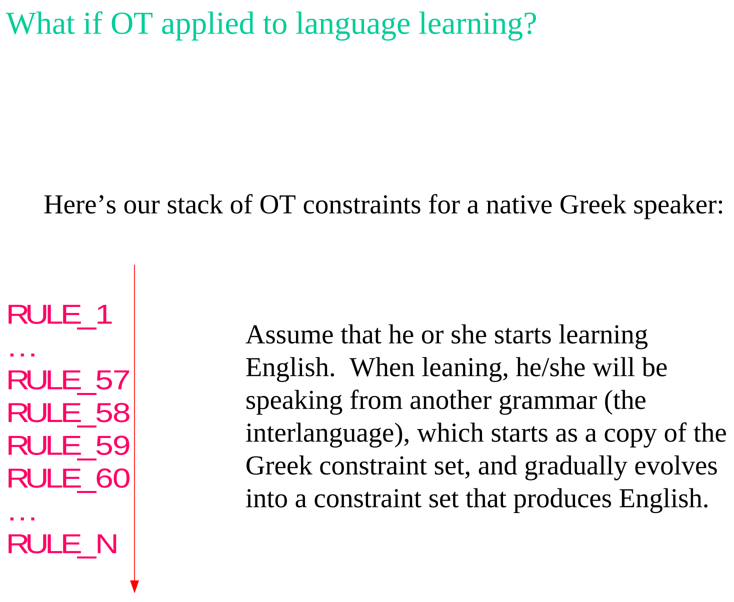
If OT is to be considered a theory, and therefore useful in a scientific sense, its only hope is to be a compact description of language. Unfortunately, no one has tried the experiment: trying to find an OT representation (ideally for all languages) that is more compact than simply listing the data that was used to create the OT representation.

Now, there has been work on OT to add extra constraints to the original framework, but I'm not aware of any modifications that would affect these arguments and make OT into a real theory.

3.2 Optimality Theory – re-Ordering by Blocks

Unfortunately, OT as it stands, is not testable because the nature of the constraints is unspecified. Figure 3 shows a straightforward example based on and idea by **?**. (Nathan pointed out that the sub-grammar of English that describes papers on OT requires the rule **+FRENCH** to convert "Table" to "Tableau".) This shows a simple Optimality Tableau for French. The top-ranked constraint, **+FRENCH**[3], ensures

[3] People who wish to stay close to the existing OT literature can consider **+FRENCH** to represent a large block of finer-grained constraints.



What if OT applied to language learning?

Here's our stack of OT constraints for a native Greek speaker:

RULE_1
...
RULE_57
RULE_58
RULE_59
RULE_60
...
RULE_N

Assume that he or she starts learning English. When leaning, he/she will be speaking from another grammar (the interlanguage), which starts as a copy of the Greek constraint set, and gradually evolves into a constraint set that produces English.

Figure 4: OT is not falsifiable.

that at most one phonological implementation for each utterance passes further down the stack.

Since only one implementation of the utterance passes down, the lower levels of the stack (**+GERMAN** and **+ENGLISH**) cannot influence which implementation is chosen; whether they pass it or not, it is still the unique best implementation. Now, we can trivially represent any language by simply re-ordering the list. By placing **+ENGLISH** above **+FRENCH**, the tableau will generate perfect English.

Therefore, if an OT representation exists for each language separately, one can re-order the list to represent any language. Few, if any papers on OT raise the question of whether or not OT *can* represent languages; that is assumed. Instead, the literature is focused on *how* to represent languages.

Again, we can see that OT is under-constrained.

4 Testing Optimality Theory in Language Learning

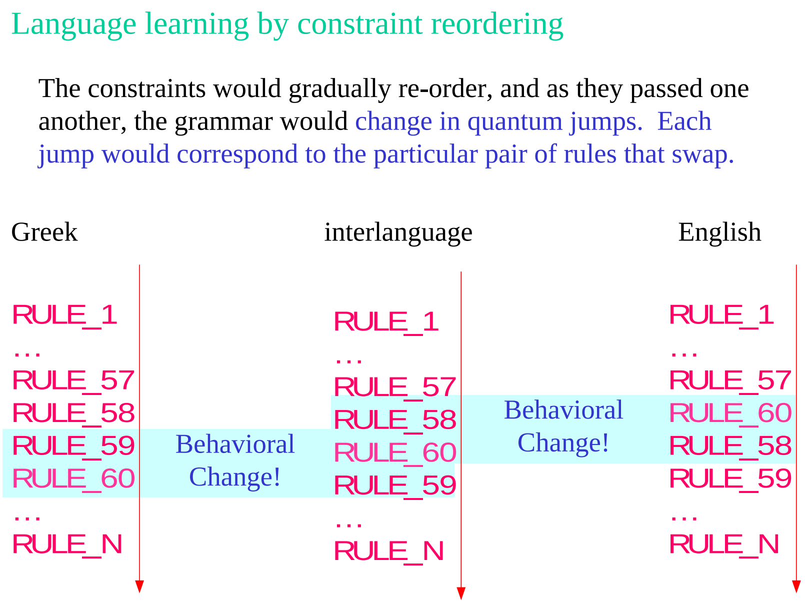
Although OT itself isn't a falsifiable theory, it could be converted into one[4]. Let's suppose that someone converted OT into a serious theory. How would we test it? Looking at the process of language learning provides an avenue for testing OT.

Language learning is sometimes described as a creation of an "interlanguage" (Figure 4, which starts out as a copy of the first language. The interlanguage is then gradually modified as the new language is learnt. Eventually, as fluency is achieved, it becomes similar to the language of a native speaker.

If we transform that description directly into the language of OT, the interlanguage begins as a copy of the main language's constraints[5] in the same order. As the new language is learnt, the order of the

[4] Whether this is a good idea or not is another question. Optimality theory is so different from the way the brain functions that it is hard to imagine that it could be correct.

[5] By definition, since all languages share the same constraints.



Language learning by constraint reordering

The constraints would gradually re-order, and as they passed one another, the grammar would change in quantum jumps. Each jump would correspond to the particular pair of rules that swap.

Greek

RULE_1
…
RULE_57
RULE_58
RULE_59
RULE_60
…
RULE_N

Behavioral Change!

interlanguage

RULE_1
…
RULE_57
RULE_58
RULE_60
RULE_59
…
RULE_N

Behavioral Change!

English

RULE_1
…
RULE_57
RULE_60
RULE_58
RULE_59
…
RULE_N

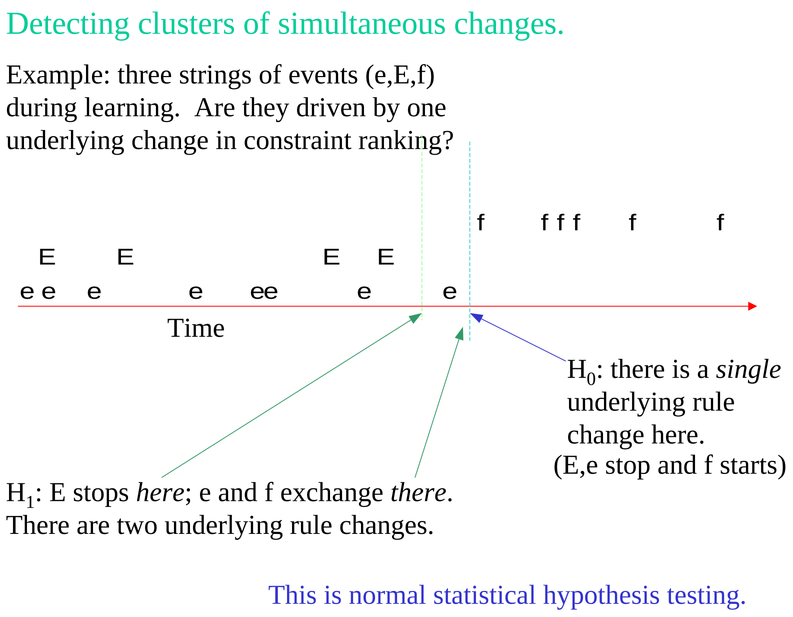
Figure 5: Constraint re-ordering in language learning.

rules gradually changes. We will assume that the constraints move smoothly: that if you were somehow able to watch the interlanguage's rule set, it would change slowly. From one minute to the next, the order of the constraints will usually be the same, with close to the minimum possible number of changes.

When a pair of constraints cross (Figure 5), the interlanguage suddenly changes. In most extant versions of optimality theory, the constraints do not correspond directly to the proper pronunciation of a single word. Instead, the constraints typically refer to things like the feature of nasality, or rules for stress placement. Consequently, one would expect that a typical rule reordering would have broad effects in the language. Pronunciation of a number of words would change, and if a syntactic change were involved, one would expect the grammatical structure of many sentences to change at once.

Consequently, if one tracked several frequently occurring linguistic events, you'd see a pattern much like Figure 6. Imagine that you were tracking many events, and three of them showed a nearly simultaneous change. These events might be something like the observation of a trilled /r/ or use of SVO (subject, verb, object) sentence structure. We imagine that we are observing a language class, rather than testing the subjects, so that the events occur randomly, beyond our control. Every time we hear a trilled /r/ (e.g. ʀ) instead of an approximant (e.g. ɹ), we mark down event **f**; every use of SVO structure is marked as event **E**, and event **e** corresponds to the use of a vowel from Language I in a certain context where the speaker should really use a nearby vowel from Language II.

Since the **e** and **E** events stop at about the same time that the **f** events start, we hypothesise that these three streams of events are related to a single cause, a single constraint re-ordering. But, can we prove it?



Detecting clusters of simultaneous changes.

Example: three strings of events (e,E,f) during learning. Are they driven by one underlying change in constraint ranking?

H_1: E stops *here*; e and f exchange *there*. There are two underlying rule changes.

H_0: there is a *single* underlying rule change here. (E,e stop and f starts)

This is normal statistical hypothesis testing.

Figure 6: Watching linguistic events near a constraint re-ordering.

5 Hypothesis Testing

We want to test the hypothesis, H_0, of a single cause for the changes in the three streams of events. An alternative hypothesis, H_1, is that there are two events, one for **E** and the other that controls **e** and **f**. Another hypothesis, H_2, might be that there are three underlying causes, one for each event stream.

H_2 is the weakest of these hypotheses. If we take it as given that each stream of events starts or ends, then H_2 is certainly true. H_0 is a subset of H_2, of course, so it's completely possible for both to be simultaneously true. (H_0 is just H_2 with the added constraint that all three transitions happen at the same moment.)

- If H_0 is true, then seemingly unrelated linguistic changes come in clusters. We will count that as support for OT, because OT (along with some plausible hypotheses about how it might operate during language learning) predicts broad changes in the interlanguage for each rule interchange.
- If H_0 is false but H_1 (or its variants) are true, then we can count this as weak support for OT. A pair of event streams seem to be controlled by the same underlying grammar change, but the other event stream, **E**, seems to be on its own.

  This extra grammar change might be bit of an embarrassment for OT, though that depends on how completely we are monitoring the subject's language. If we were only to monitor a few properties, then it would be easy to imagine that event stream **E** was paired with some other sequence of events that we are simply not watching. In that case, H_1 remains completely consistent with OT.

  Likewise, if there are phonological properties that we monitor, but which simply aren't seen very often, it might not be a problem. **E** might be paired with some observed but rare property that only occurs once a week (e.g. stress clash in a double-embedded subjunctive construction), and it simply didn't happen near the **e**→**f** transition.

On the other hand, absent some plausible excuse for pairing the **E** stream with an unobserved set of events, such a single, isolated change would be a rather unexpected result of a typical constraint interchange in an OT model of a language.

We will later estimate how much evidence H_1 would provide for or against OT, depending on various assumptions.

- Finally, H_2 is best interpreted as evidence against OT. We would have looked for an expected result (multiple correlated changes), and found the reverse. We will later estimate how much evidence H_2 would against OT. Would once be enough to disprove OT?

But first, we will look into hypothesis testing, to see how we would decide whether H_0, H_1, or H_2 might be true.

References