

EXPERIMENTAL SYNTAX - AN INTERVIEW WITH THOMAS ROEPER

Thomas Roeper

University of Massachusetts

REVEL - What is Experimental Syntax? How does it relate to Theoretical Syntax and how does it relate do Psycholinguistic studies?

TOM ROEPER - Experimental syntax is a big field. It certainly encompasses all of the work that links specific hypotheses from linguistic theory to psychological experiments that engage those structures.

The term is fairly new, though, and I want to give it a particular twist here which not all psycholinguists would agree with. One way to view the world of psycholinguistics is to say that there is “theory” on one side, with obscure intuitions to back it up, and the world of established, well-defined methodology from psychology on the other. One simply brings them together.

In my view, once one accepts the core claims of modern grammar, it provides a totally new perspective on all methodologies--seizing some parts and downplaying others. The critical idea is that we do not--probably cannot--perfect methodology by eliminating all potential confounding factors, but we can seek independent confirmation which is far more powerful than eliminating all experimental variables. Independent confirmation is the right concept, because linguistic theory is already grounded in grammaticality intuitions. It carries its own empirical support in the form of grammaticality intuitions.

Often that is sufficient. We don't need to do an experiment to see that *hat the is ungrammatical in English, but not the hat. However many grammaticality judgments come with flags--a question-marker--indicating some uncertainty.

It is still important empirical evidence, but it invites independent confirmation. Andrew Weir (pc) has observed that there is a subtle difference between two diary entries, each perhaps questionable in typical environments:

- a. found phone on the street today = random phone
- b. I found phone on the street today = my phone

Now a difference like this--which is fairly evident but subtle--might receive further support if we developed a psycholinguistic experiment to measure reaction time or an acquisition scenario to show that 5 year olds would make the same difference. Suppose we did that, but each experiment has various imperfections, or flaws, but fits the basic account. What would that mean and how seriously should we take it to be? I would argue that yet another more refined experiment might not be warranted. The very fact that we get any results that fit such a subtle contrast should be taken as strong evidence for the claim because it comes from two different sources. It is the same logic that we find when we see weak subjacency effects in several constructions, but they all are found across languages and fit one theoretical account of barrier nodes.

This is the real method of experimental syntax--a quite different method from simply importing constraints from psychology that are efforts to define methods independent of theory.

Take another case. Someone once told me that all experiments with children had to have equal numbers of boys and girls at each age--good methodology, supposedly, but experimentally demanding. Why exactly should that be? It is possible that one could make different hypotheses about boys and girls. That is true, but until such a hypothesis exists, it is not at all clear that it is necessary. Ultimately all experimental methods are mini-theories which need to be justified if we are to take them seriously. If we do not have such justification, then we should adhere to the logic of

“independent confirmation” which almost any evidence deriving from an experiment that agrees with a pre-existing theory, provides.

By the way, this is the standard in other sciences. Einstein made very subtle predictions about planetary behavior on the basis of relativistic physics. Much of the data is thrown out from photographs because it does not show something one way or the other. If one finds one piece of data that fits, then the physicists take it very seriously--precisely because it conforms in detail to an independently conceived theory. That should be our practice as well.

I consider many acquisition experiments to be comparable to “field linguistics” where only small amounts of data are readily obtainable. Yet if they conform to a sophisticated theory, then we take them seriously. The same holds for taking acquisition experiments into the field where we may just have a few living speakers for some language. If we can obtain informal results from a few speakers that conform to linguistic theory, established elsewhere, then we should take it very seriously.

It is really important that journals realize the power of this logic and do not apply methodological restrictions that block a wider audience from seeing important results. Long ago, I did an experiment with Larry Solan that showed that children never misattached a relative clause if it is inside a VP. Here is how it went. We compared these sentences:

- a. the rat pushed the cat that the dog hit
- b. the rat put the cat that the dog hit in the barn

In (a) children will often allow “that the dog hit” to apply to the upper NP, the rat. In (b), the structure is different: [put [relative clause] in the barn] the particle verb VP is outside the relative clause and it would be necessary to cross-lines--totally disallowed in linguistic theory (with a misleading exception in multi-dominance theory) for good empirically established reasons. The results were 100% in obedience to what linguistic theory predicted: no children thought that the rat hit the dog in (b) where 30% did in (a). It is hard to imagine that this is not convincing evidence when it

brings together both a very sophisticated theory and completely independent acquisition evidence. But there were only 17 children, not 20 in the experiment and a journal said it failed to meet the statistical criterion. Such a response fails to see that two independent forms of evidence are converging. That is what we should aim for, not one perfect experiment.

Now if a field linguist were to do the same experiment with four speakers of an unusual language and obtain the same results, such results are also extremely important because they bring a language from a different family in support of the hypothesis. It would be crazy to disallow the evidence because they did not have 20 speakers. And it would be very unfortunate if the research is not funded in the first place because laboratory conditions cannot be reproduced. This kind of converging evidence is, I believe, the essence of work in linguistics and applies to the entire enterprise.

REVEL - You have been doing experiments to tap knowledge of language for many years. Traditionally, the main and sometimes the only methodology used in Generative Grammar used to be the elicitation of grammaticality judgments. In the last few years, though, many theoretical linguists have turned their attention to experimental methodologies giving rise to the field of “Experimental Syntax”. How do you see this new field and how do you compare grammaticality judgments to the fancy experiments which are being carried out these days? How do you compare acceptability judgments to grammaticality judgments?

TOM ROEPER - The answer here follows from my previous remarks. I actually do not believe that we should create an abstract comparison of such techniques--it is not possible nor advisable. What you do is see how well any results lead to the support or non-support of a given theory. It is widely believed--too strongly in my view--that no psycholinguistic results can challenge the conclusions arrived at via grammaticality intuitions. Such intuitions, it is supposed, go directly to grammatical representations without performance interference. But, first, we know that there are many kinds of

semantic, pragmatic, and processing forms of interference in such judgments. Second, we have “??” judgments which are weaker than “ungrammatical”. They seem to mean “maybe grammatical” but no one is sure why they are weaker. Now we really have to attend to other kinds and sources of evidence. If such constructions are examined with totally different techniques, then we should let the other psycholinguistic evidence determine whether they should be treated as grammatical or not. Even if we accept them as grammatical, we still need to explain where the “maybe” came from. Here again what look like interfering factors need to be elevated to the status of theories by seeing if we can formulate what factor is causing these effects.

Ultimately, we can only compare theories. All of the “methodological factors” are really alternative theories which need to be articulated. Suppose we make the commonsense and practical observation that children “don’t pay attention” to long sentences. Now we need to explore that in theoretical terms. If I say “very, very, very, very strong” the sentence is getting longer, but the processing demand is not because we have a technique for representing iteration that is different. The same logic holds for conjoined versus embedded relative clauses. Some children string along conjoined clauses at great length, something like: “that’s the train I got for Christmas, that my dad gave me, that I really like, that goes really fast.”

The conjunction makes it easier, and maybe children can “pay attention” more easily too. So now, whatever “attention” is, we will find ultimately that it defines a subtle interface with grammatical representations that we should understand. I like seeing new techniques come into the field and I believe that they can make an important contribution. However, without a clear vision of what scientific affirmation is, the so-called methodological factors can impede progress. This is actually the position that Chomsky originally took against some of the empirical structuralists who believed that mental phenomena, including intuitions, were illusions, and linguistics should only deal with spoken sentences, not intuitions.

The consequence was the triumph of the opposite view: only intuitions really tell us about grammar. It is an over-reaction too. In sum, the notion that methodologies are independent of theories is really harmful. Discussions about “validating

methodology” builds edifices which are never scrutinized once they are established. Chomsky once remarked that “one should use whatever evidence one can find”. Consider, for instance, anecdotal evidence in language acquisition. It has been criticized because parents are biased and misremember. That is true. As someone who is both a parent and a researcher, though, I know how much I learn from anecdotal and naturalistic evidence in language acquisition. I always respect it, and never take it as definitive.

Once you hear something suggestive, you can go out and create an experiment to test it. Andrea Faber recently noted that her daughter said “these dolls are taller than each other” which is an ungrammatical sentence from an adult perspective (though it has been suggested that it can occur in French). We know that “the chairs are on top of each other” is acceptable, though it seems not to make sense and provides a substantial challenge to semanticists. So the next step is to create controlled experiments and see if other children accept such sentences and how they interpret them.

The entire Childe database--which has proven to be enormously fruitful--is equivalent to a poorly controlled experiment because the context is not controlled and not represented in the database, nor is the phonology present most of the time. So all of it calls for further investigation. In particular we can see children using a construction, but we do not know if they control all of its uses.

A great deal of work, currently--a number of dissertations--looks at the logical properties of “only” which was first investigated in acquisition by me, Steve Crain, Bill Philip, and Ken Drozd in 1992 where we found that sentences like: only John has an umbrella \ John has only an umbrella

both allowed both readings: “only John” and “only umbrella”. If one goes into the database one finds things like: “Only I want milk” with the meaning quite obviously of “I want only milk”.

And yet a deeper question is lurking here which Barbara Höhle and I and her students in Potsdam are pursuing: does only apply to the whole sentence and allow

event contrasts? Most parents have experienced a conversation like this with a 4 or 5yr old:

Mom: “you have to go to bed”
“only I want to see the end of the movie”

This is not “only I” or “only movie”, it is really a comment on the previous clause (go to bed) with another clause (see the end of the movie). This may be the first form of “only” that children have and maybe we have misread the database case “only I want milk”. So this calls for independent experimentation.

And it provides a sharp contrast to the prevailing research preferences. Nevertheless, not only for children but for adults such “conjunctive” uses of “only” are important. So my answer has wandered far afield from the question of comparing fancy methodologies. I hope, though, you see the relevance of concrete reasoning with a theoretical backdrop, for all such questions.

REVEL - Tom, in a paper you presented last year in Brazil’s *First International Congress of Psycholinguistics*, in Rio de Janeiro, you mentioned the concepts of “slow thought” and “fast thought”. Could you elaborate on them?

TOM ROEPER - The contrast does not really originate with me. Chomsky on a variety of occasions has distinguished between rapid mental processing in milliseconds and slower problem-solving behavior. An extreme example is the difference between doing an implicit calculus calculation when throwing a basketball--which is fast--and solving the same problem with pencil and pen, which is slow. Moreover we intuitively think that these two types of computation--again, in fact, solving the same problem--are solved in totally different ways. And, of course, you can be good at either without being good at the other. We have other forms of fast and slow thought and there is a whole new book on the topic *Thinking Fast and Thinking slow* by Daniel Kahneman. I think we have more to learn by looking at language closely than via the examples that he gives. Yet let us consider the extra-linguistic domain for a moment.

If I contemplate over several weeks whether I want a new computer, a trip to Hawaii, or to remodel my house, I am engaging in thinking that I want to be slow and that involves assessing inherently incommensurate information. The manner in which we do that is utterly unknown. One has a hunch, though, that we change our presuppositions in that process. We make the assumption that number one is an immediate vacation, or number one is fixing a decaying house, and our decision depends upon which way we set up the priorities, which seems to take time to realign. So we can see that presuppositions behind thinking and changing them is important.

Now those observations are really step one. If we push the idea further, it raises other questions. Let me start with another example built on extremes. The Shakespearean line “ripeness is all” can be understood in milliseconds like any sentence by a 16 yr old. But one might say “you have to be over 60” to really understand what the phrase means. Well, that is a big difference: 100 milliseconds and 40 years. Yet, ultimately, we do need to characterize both forms of sentence comprehension.

The issue becomes more acute when we move to sentence-processing and conditions for acquisition. We process sentences very quickly, but we also compute background pragmatics and the determination of topics, presuppositions, and the question-at-issue. Some of this is very fast, but is it all fast?

Let me make up some numbers since I don't have parsing facts at my fingertips. Suppose I observe that most sentences are understood in 100 milliseconds, with variations for length and complexity that go from 50 to 150 for complex sentences like: what did the man that the girl saw say___

Now compare that with: John's son is in a race. John believes that Bill, unfortunately, will win the race.

Suppose--and I am not sure I have the best example here--that it takes 1000 milliseconds--a full second, to give a response to that sentence via some time-measurement technique. It is not long, but it is 10 times longer than the original

parse of a more complex sentence. Why would it be so long? What kind of computation of the “unfortunately” is involved? It seems to be there to capture the implication that it is unfortunate from John’s point of view (not Bill’s) that Bill will win the race because, plausibly, John hopes his son will win the race. That is not just background but a set of consequences, logical and social, which are engaged by the word “unfortunately”. It belongs in our broader semantics because we are not understanding this sentence unless we compute it. And yet its status as a mental activity seems unclear, much slower, and perhaps belongs to an almost different type of thinking (which is what Kahneman says, although his version of slow thought would not get down to a difference between 100 and 1000 milliseconds).

This kind of computation can be stretched out too. Consider a conversation with someone who says “my friends will help me” and the remark seems innocuous. Yet a few days later you meet someone else who notably says “I won’t help him”. And then suddenly you realize that he really said “my FRIENDS will help me” with a contrastive implication that “my ENEMIES or non-friends” will not, and then you realize that the statement was fearful or soliciting assistance. This is slow, but a related computation of the implications of a single sentence is still involved, even though it occurs a few days later. So the slow thought style of reasoning may be on a continuum with some fairly fast thought phenomena.

A lot of current psycholinguistic work is relevant to these questions. There is a great deal of exciting work on pragmatics, implicatures, and how the Common Ground is computed. There are interesting suggestions that autistic children do not manage it very well. There are claims that children cannot compute implicatures until they are 6 years old, and other claims that they can do so much earlier.

And there is interesting work by Anna Verbuk on conventional implicatures, which form a particularly interesting subclass of implicatures. She gave children sentences like: The fisherman told the visitors that there were no fish in the pond, and then we went and fished there. Honestly, the fisherman lied.

And when she asked 6yr olds if the fisherman was honest, many said yes, while 8yr olds said no. What's the challenge here? One must realize that the "honestly" reflected the perspective of the speaker and not the subject.

We manage this very rapidly, but yet as I suggested, it may be related to the other slow thought skills that are culturally dependent, that we typically undertake over longer periods of time and that may reflect many individual differences.

In sum, slow thought remains quite mysterious. It seems to involve things like shifting presuppositions and it clearly relates to the kinds of thinking we do consciously when we gradually realize that there is a hidden contradiction between two sentences.

In order to understand how this dimension of language works, we need to conceptualize rather broadly the kinds of mental activity that are involved. It is a big task and it will not succumb quickly to inquiry. We are beginning to build the interface between processing and thought in general--which can be very slow--and yet that domain of slow thought continues to be probably the greatest scientific challenge that exists. No question in biology is more sophisticated than how the notion of how mind arises in human beings and in some form in many other species. It is both gratifying and humbling to see the processing and acquisition of language begin to take on those questions.

REVEL - What are some hot topics for experimental syntacticians today?

TOM ROEPER - A number of classic questions have occupied both linguists and psycholinguists for a quarter of a century. They include how people comprehend barriers and various forms of embedding. More recently the question of how to represent recursion has become important in linguistic theory, acquisition and now fieldwork. In addition the challenge of how quantification works remains an important topic in semantics with experimental data playing an ever larger role. And a natural connection between pragmatics and experimental methods has been recognized more widely (there is now a conference on experimental pragmatics).

What this means is that we are beginning to grasp theoretically the pragmatic setup behind experiments, just what we called “methods” from the perspective I advocated above. That is, all methodology is really theory in disguise. Every methodological practice should be examined and elevated to a theoretical option which requires its own rigorous analysis and defense. Concepts in psychology that seem like rock-bottom assumptions, such as “attention”, interact with such highly complex ways with all mental structures that we need to subdivide and dissolve them. As I said above, if a theory of biology could arrive at a conception of gender differences which lead to really different acquisition or processing differences--which I strongly doubt--then we could consider them.

REVEL - Could you please suggest some essential readings on Experimental Syntax for our readers?

TOM ROEPER - Well let me not comment about literature in parsing, with which I have some familiarity, but I would not hazard a recommendation. And in acquisition, I would like to point to the *Handbook of Generative Approaches to Acquisition* (Springer) which Jill deVilliers and I edited and appeared in Fall 2011. It has, we believe, a good overview of the field from early to recent work. We hope it will inspire a lot more work.

Thank you.