★ FORUM ★

Continuous Acceptability, Categorical Grammaticality, and Experimental Syntax

Jon Sprouse

1. Introduction

It almost goes without saying that acceptability judgments form a continuous spectrum. While many sentences are either clearly acceptable or clearly unacceptable, a significant number of sentences fall somewhere in between in a gray area of partial acceptability. This fact has been explicitly admitted by linguists since at least Chomsky 1965. The working assumption adopted by most linguists over the past 40 years has been that these intermediate levels of acceptability are caused by properties other than grammatical knowledge. Linguists have assumed that grammatical knowledge is categorical — sentences are either grammatical or ungrammatical — and that the continuous spectrum of acceptability is caused by extra-grammatical factors (plausibility, working memory limitations, etc.). Of course, ideas such as *strength of violation* have been introduced into theories of grammatical knowledge at various times, for instance Huang's (1982) proposal that ECP violations are *stronger than* Subjacency violations or Chomsky's (1986) proposal that each barrier crossed leads to lower acceptability. However, with the notable exception of Optimality Theory (see especially Keller 2000, 2003), these analyses have been the exception rather than the rule.

The past ten years or so have seen a major shift in attitudes toward intermediate levels of acceptability. With the increasing use of formal experimental methods for measuring acceptability — a methodology that has come to be known as experimental syntax — it has become almost trivial to detect subtle differences along a continuous spectrum of acceptability. This new power afforded by experimental syntax raises the question of whether the continuity of acceptability reflects a continuity in grammatical knowledge that should be captured by the theory of grammar, or in other words, whether the working assumption of the past 40 years should be abandoned (see especially Keller 2000, Fanselow *et al.* 2004).

While the answer to this question is ultimately an empirical one that is far from being settled, this report presents two pieces of experimental evidence for a categorical distinction between grammatical and ungrammatical sentences. The first is a direct prediction of theories that assume categorical grammaticality. The psychological claim underlying theories of categorical grammaticality is that ungrammatical sentences have no licit representation, or in other words, cannot

biolinguistics

ISSN 1450–3417
Biolinguistics 1: 123-134, 2007
http://www.biolinguistics.eu

be constructed from the available mental computations. Grammatical sentences, on the other hand, have licit representations that can be constructed from the available mental computations. This predicts that extra-grammatical factors that affect the acceptability and are predicated on the existence of a representation, such as syntactic priming (Luka & Barsalou 2005), should not affect the acceptability of ungrammatical sentences. Section 2 presents results from Sprouse (2007) that confirm this prediction: Ungrammatical sentences, in particular island violations (Ross 1967), do not show a structural priming effect.

The second piece of evidence comes from the experimental syntax technique magnitude estimation. Unlike traditional tasks such as the *yes*/*no*–task and the Likert scale task in which subjects must categorize their responses, magnitude estimation allows subjects to respond using the theoretically infinite continuum of values available on the positive number line (see Bard *et al.* 1996). By removing the categorization aspect of the task, one might expect that responses would no longer show any categorical distinction between grammatical and ungrammatical sentences. Contrary to this prediction, section 3 presents evidence from Sprouse (2007) that subjects impose a form of categorization on magnitude estimation responses, and that this categorization appears to correspond with the grammatical/ungrammatical distinction.

2. Syntactic Priming and Categorical Grammaticality

Syntactic priming is the facilitation of a structure through previous exposure to that structure. For instance, speakers tend to repeat structures that they have either heard or spoken recently (e.g., Bock 1986, Pickering & Branigan 1998), and readers tend to show faster reading times for structures that they have recently read (e.g., Noppeney & Price 2004, Kaschak & Glenberg 2004). Of particular interest to our present purposes is that Luka & Barsalou (2005) found that exposure to structures in a reading task increase the acceptability of those structures in a subsequent rating task. This suggests that syntactic priming can be indexed by acceptability rating tasks, or to put it another way, that acceptability ratings are affected by repetition.

Syntactic priming is the effect of one sentence on a structurally identical subsequent sentence. This entails the *possibility* of constructing the structural representation in question. From the perspective of theories of categorical grammaticality, only grammatical sentences have representations that can be constructed, therefore only grammatical sentences should show syntactic priming effects. While Luka & Barsalou (2005) demonstrate a syntactic priming effect for (by hypothesis) grammatical structures, ungrammatical structures were precluded from the priming analysis because ungrammatical sentences could not be presented to the subjects during the reading phase. Presentation of ungrammatical sentences during the reading phase could have drawn attention to the dimension of acceptability, and potentially biased the subjects prior to the acceptability rating task.

The idea of syntactic priming of acceptability judgments is well known by syntacticians concerned with acceptability judgments. It has long been anec-

dotally reported that judgments tend to get better over time, a phenomenon that Snyder (2000) calls *syntactic satiation*. In fact, Snyder presents evidence that naïve subjects show a satiation effect with two ungrammatical structures (*wh*-islands and Complex NP Constraint islands) in a *yes*/*no*–acceptability task. Prima facie, this appears to be evidence for a syntactic priming effect of ungrammatical sentences, contrary to the predictions of theories of categorical grammaticality.

Sprouse (2007) re-examines Snyder's evidence for satiation in detail. In that paper, I argue that Snyder's design introduces a confound that may be responsible for his results: Subjects may adopt a strategy in which they attempt to equalize the number of *yes*– and *no*–responses, resulting in an increase in *yes*–responses over the course of the experiment. While the evidence for this confound cannot be reviewed here (the reader is referred to Sprouse 2007 for details), the re-designed experiments that rectify this confound can. As will become apparent shortly, once the *yes*/*no*–strategy confound is eliminated, there is no evidence for syntactic priming of ungrammatical structures in acceptability judgment tasks, as predicted by theories of categorical grammaticality.

2.1. Rationale and Design

Theories of categorical grammaticality predict that ungrammatical sentences will not be affected by syntactic priming. To test this prediction, four experiments were conducted, one experiment for each of the following island violations:

(1) *Islands tested for syntactic priming (satiation)*

Subject island:	Who do you think the email from is on the computer?
Adjunct island:	Who did you leave the party because Mary kissed?
Wh-island:	Who do you wonder whether Susan met?
CNPC island:	Who did you hear the rumor that David likes?

Over the course of the Subject and Adjunct island experiments, subjects were exposed to 14 instances of the island violation. Over the course of the *wh*- and CNPC island experiments, subjects were exposed to 10 instances of the island violation.

Two design features were incorporated in these experiments in order to eliminate the possibility of the *yes*/*no*–strategy. First, the task was magnitude estimation. The theoretically unlimited number of levels of acceptability made possible by magnitude estimation decreases the possibility of subjects counting the number of each type of response. Second, the number of grammatical and ungrammatical sentences were balanced in each experiment. Even if subjects attempt to track the number of responses (perhaps with respect to the modulus sentence; see section 3 below), the distribution is already balanced, so no change should result. Judgments were collected for each repetition of the island violation. A syntactic priming (satiation) effect would result in an upward trend in acceptability, which could be confirmed statistically through linear regression. All of the subjects were undergraduates at the University of Maryland with no formal training in linguistics. The sample sizes for the experiments were 20, 24, 20, and 17, respectively.

2.2. *Results*

Following the standard data analysis procedures in Bard *et al.* 1996 and Keller 2000, responses were divided by the reference sentence and log transformed prior to analysis.[1] The means for each exposure are given in the graphs below.

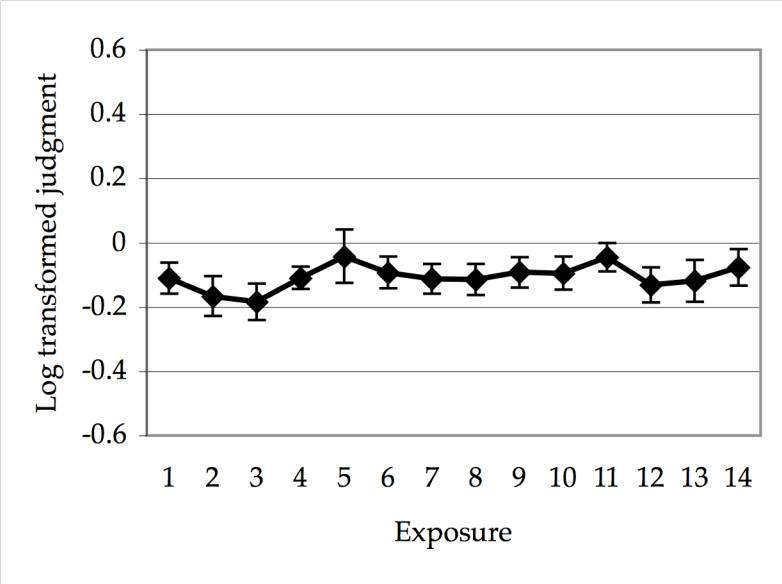

Figure 1: Subject islands

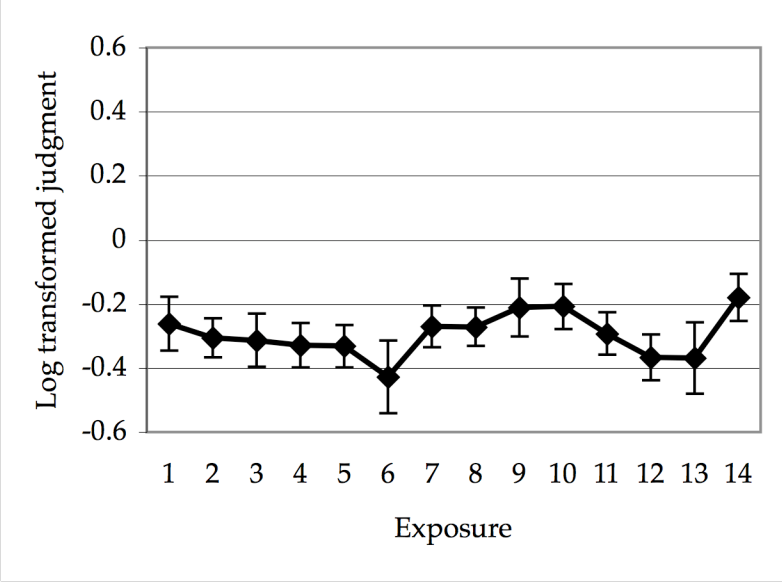

Figure 2: Adjunct islands

[1] See Sprouse 2007 for evidence that the log transformation may be unnecessary.

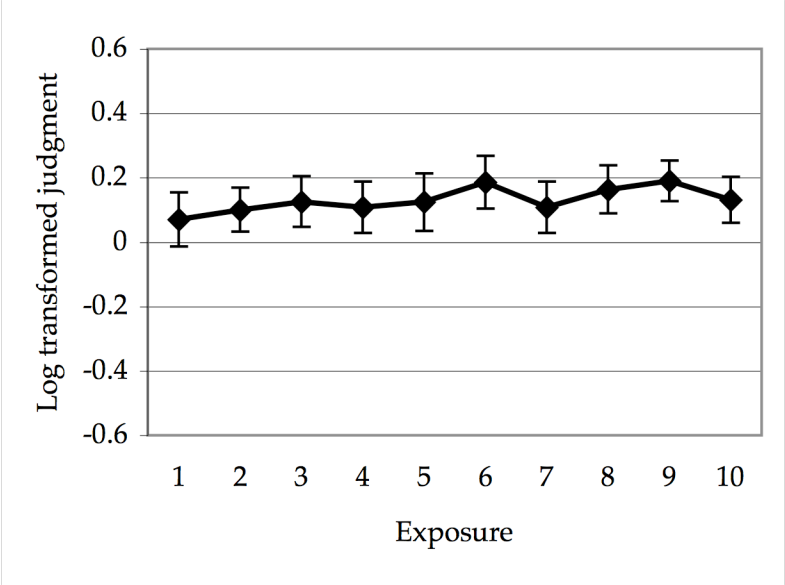

Figure 3: *Wh*-islands

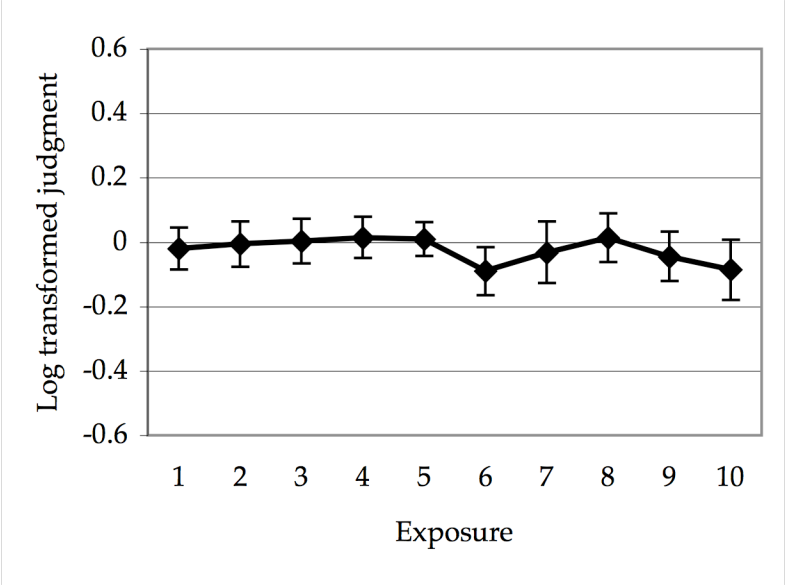

Figure 4: CNPC islands

Repeated measures linear regressions (Lorch & Meyers 1990) confirm that there are no significant increases in acceptability for any of the islands tested.

	b	*Exp(b)*	p
Subject island	–0.13	0.003	.14
Adjunct island	–0.32	0.003	.52
Wh-island	0.08	0.008	.14
CNPC island	0.001	–0.006	.44

Table 1: Linear regression of means

2.3. Discussion

Categorical grammaticality predicts that some extra-grammatical effects on acceptability may be asymmetrical, affecting grammatical structures but not ungrammatical structures. Syntactic priming is one such factor: The priming effect is predicated upon the existence of a licit representation. Given that ungrammatical structures have no licit representation, categorical grammaticality predicts that there should be no syntactic priming effect for ungrammatical structures. While Luka & Barsalou (2005) demonstrates syntactic priming effects on acceptability for grammatical structures, the results of these experiments suggest that there are no syntactic priming effects on acceptability for ungrammatical structures, as predicted by categorical grammaticality. These results also suggest that asymmetric extra-grammatical effects on acceptability may be a useful tool for identifying the grammatical status of structures that are neither clearly grammatical or clearly ungrammatical.

3. Magnitude Estimation and Categorical Grammaticality

In many ways, the magnitude estimation task has become synonymous with experimental syntax. The idea behind the task is simple, and perhaps easiest to explain with an example: Imagine that you are presented with a set of lines. The first line is the *modulus* or *reference*, and you are told that its length is 100 units. You can use this information to estimate the length of the other lines using your perception of visual length. For instance, if you believe that the line labeled item 1 below is twice as long as the reference line, you could assign it a length of 200 units. If you believe item 2 is only half as long as the reference line, you could assign it a length of 50 units. Figure 5 illustrates:

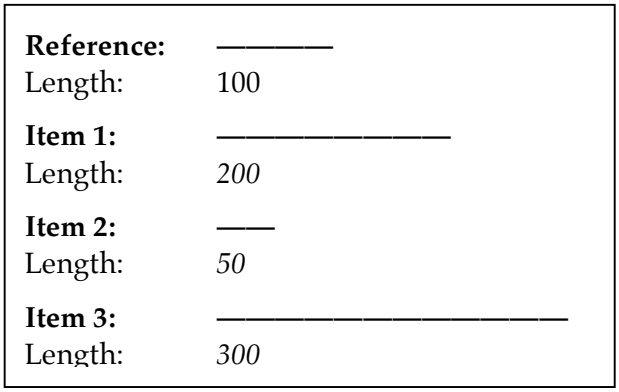

Figure 5: Magnitude estimation task

The resulting data are estimates of the length of the items in units equal to the length of the reference line.

Bard *et al.* (1996) were the first to argue that this technique could be adapted for the estimation of acceptability. For instance, imagine you are presented with a pair of sentences. The first is the reference sentence, and you are told that its acceptability is 100 units. The acceptability of the second sentence can then be estimated using the acceptability of the first as a reference:

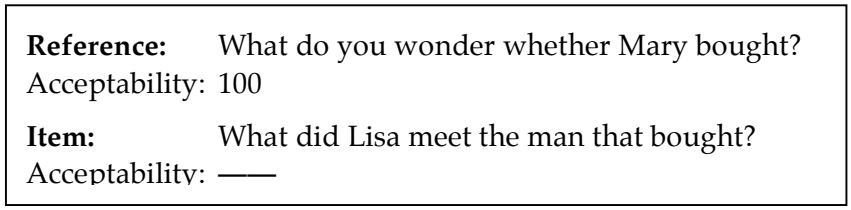

Figure 6: Estimation of acceptability

Bard *et al.* argue that this task is superior to the standard scale tasks usually used in acceptability studies because (i) it allows subjects to distinguish as many levels of acceptability as they want, unlike scale tasks in which they are not limited to 5 or 7 choices, and(ii) the distance between levels of acceptability are measured in regular units (equal to the acceptability of the reference sentence), unlike scale tasks in which the distance between points is lost to categorization.

The two benefits of magnitude estimation suggested by Bard *et al.* have in many ways become a catalyst for considering continuous models of grammatical knowledge. The freedom to distinguish a theoretically infinite number of levels of acceptability and the ability to quantify the distance between those levels with a standard unit are exactly what is needed to precisely model continuous acceptability. However, it is the underlying assumption of linearity that truly enables a continuous model of grammaticality: The magnitude estimation task makes no explicit distinction between grammatical and ungrammatical

structures. They are both measured using the same reference (measured in the same units), as if they form a linear system. Sprouse (2007) investigated the linearity assumption of magnitude estimation data in detail. As will become evident below, one of the results of that study suggests that subjects introducing a categorical distinction in magnitude estimation data that appears to mirror the theoretical distinction between grammatical and ungrammatical sentences.

3.1. Rationale and Design

Two magnitude estimation experiments were conducted, each identical in content and design, except for the reference sentence. The reference sentence for the first experiment was a type of *wh*-island violation, specifically an *if*-island: *What do you wonder if Larry bought?*. The reference for the second experiment was a Coordinate Structure Constraint violation: *What do you think Larry bought a shirt and?*. The critical manipulation is that the reference sentence for experiment 2, the Coordinate Structure Constraint (CSC) violation, is one of the conditions in the experiment. Therefore the values obtained from the first experiment can be used to predict the values in the second experiment. We know that the value of the CSC in the second experiment is going to be 1 (because it is the reference unit). So for instance, if the CSC is rated as .5 in the first experiment, we would expect all of the values in the second experiment to be doubled (because .5 x 2 is 1). If the CSC is rated as .25 in the first experiment, then we would expect all of the values in the second experiment to be quadrupled (because .25 x 4 is 1). In other words, the distribution of the results will remain constant, but the absolute values of the results will be transformed by the value necessary to translate the CSC into a single unit.

The body of the experiments were identical. Each contained 8 different violations types:

(2) *Violations judged in magnitude estimation experiments*

a. *Adjunct Island*
What does Jeff do the housework because Cindy injured?

b. *Coordinate Structure Constraint*
What did Sarah claim she wrote the article and?

c. *Infinitival Sentential Subject Island*
What will to admit in public be easier someday?

d. *Left Branch Condition*
How much did Mary saw that you earned money?

e. *Relative Clause Island*
What did Sarah meet the mechanic who fixed quickly?

f. *Sentential Subject Island*
What does that you bought anger the other students?

g. *Complex NP Constraint*
What did you doubt the claim that Jesse invented?

h. Whether-*Island*
What do you wonder whether Sharon spilled by accident?

Subjects judged five of each violation type. All subjects were University of Maryland undergraduates with no formal training in linguistics. 22 subjects participated in the first experiment (*If*-reference), 31 participated in the second (CSC-reference).

3.2. Results

Following the standard data analysis procedures in Bard *et al.* 1996 and Keller 2000, responses were divided by the reference sentence and log transformed prior to analysis. The mean judgments for each violation type for both experiments are provided in the following graph:

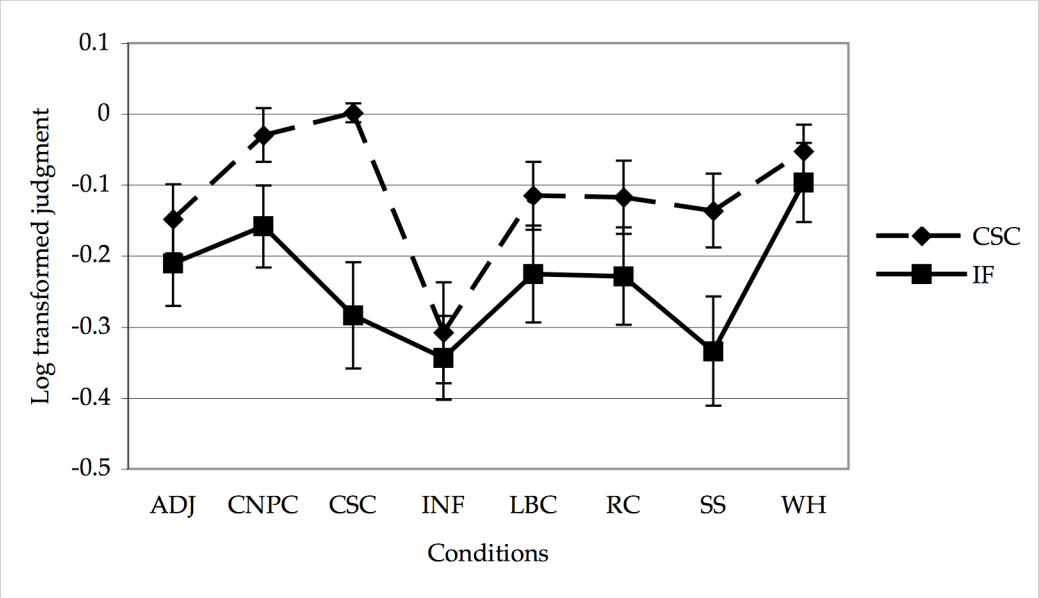

Figure 7: Mean ratings for *If* and CSC references

The first thing to notice is that the shape of the two lines is nearly identical. This suggests that the relative judgments remain constant across the two experiments. This can be quickly confirmed by comparing pairs of conditions in the two experiments. As the table below indicates, three pairs of conditions that are statistically different in the first experiment are also statistically different in the second experiment:

	If-*reference*		*CSC-reference*	
Contrast	t	*p*	t	*p*
ADJ-INF	3.86	.001	4.50	.001
CNPC-SS	3.58	.001	3.10	.002
Wh-RC	4.88	.001	1.70	.05

Table 2: t-Tests for *if* and CSC references

Therefore we can be confident that the two experiments have a similar level of sensitivity to acceptability.

Despite the high degree of agreement with respect to relative judgments, there is an anomaly with respect to the overall distribution of the judgments. Notice that the CSC violation is approximately in the middle of the distribution of judgments in the first (*if*-reference) experiment: The CNPC and *wh*-conditions are more acceptable than the CSC, the INF condition is less acceptable, and the rest are statistically equal. We would expect a similar distribution of acceptability in experiment 2, with the only difference being that the CSC should be equal to 0. However, what we find is that there are no conditions judged more acceptable than the CSC (none are above 0). Furthermore, only two are statistically equal to the CSC, the CNPC and *wh*-conditions, which were the two that were *more acceptable* than the CSC in experiment 1. The rest of the conditions are all *less acceptable* than the CSC. This is expected for the INF condition, but the rest of the conditions were statistically equal to the CSC in experiment 1

These results suggest that subjects are not actually performing the magnitude estimation task, but rather performing a relative rating task in which the reference sentence serves as an upper bound for ungrammatical items. This is also true of the results from experiment 1. In fact, in both experiments, the only conditions that are rated above the reference item are the grammatical fillers. This suggests that subjects are imposing their own categorization on an otherwise continuous task.

3.3. Discussion

By design, the magnitude estimation task provides subjects with a theoretically continuous (and infinite) response scale. As such, it is no more surprising that magnitude estimation tasks yield continuous measures of acceptability than it is that *yes*/*no*–tasks yield categorical measures of acceptability. Yet despite the continuous nature of the magnitude estimation response scale, and despite the absence of any mention of a categorical grammaticality distinction in the magnitude estimation instructions, the subjects in these experiments appear to be imposing a categorical distinction on their acceptability judgments. While the spontaneous imposition of a categorical distinction on a continuous rating scale is surprising, it is entirely consistent with a categorical approach to grammaticality.

4. Conclusion

The tools of experimental syntax have made it possible to quantify continuous acceptability with relative ease. The question this raises is whether the tools of experimental syntax have also made it possible to investigate the predictions of theories of categorical grammaticality.

This report has presented two case studies that provide support for categorical grammaticality. The first study investigated one possible prediction of categorical grammaticality: The acceptability of grammatical and ungrammatical sentences should be affected by different factors. As predicted, the study found no syntactic priming effect for ungrammatical structures, despite evidence in the literature for syntactic priming of grammatical sentences (e.g., Luka & Barsalou 2005). The second study investigated the nature of acceptability data itself, presenting evidence that subjects will impose categorical distinctions on continuous rating scales.

While these studies can only begin to address the question of the nature of grammatical knowledge, they raise the possibility of finding more evidence for categorical grammaticality in the data being collected by experimental syntax. These results argue for a closer investigation of the tasks of experimental syntax, as well as a reconsideration of the psychological consequences of categorical grammaticality in studies of the factors affecting acceptability.

References

Bard, Ellen G., Dan Robertson & Antonella Sorace. 1996. Magnitude estimation of linguistic acceptability. *Language* 72(1), 32-68.

Bock, J. Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18(3), 355-387.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1986. *Barriers* (Linguistic Inquiry Monograph 13). Cambridge, MA: MIT Press.

Fanselow, Gisbert, Caroline Féry, Matthias Schlesewsky & Ralf Vogel. 2004. *Gradience in Grammar: Generative Perspectives*. Oxford: Oxford University Press.

Huang, C.-T. James. 1982. Logical relations in Chinese and the theory of grammar. Cambridge, MA: MIT dissertation.

Kaschak, Michael P. & Arthur M. Glenberg. 2004. The construction needs learned. *Journal of Experimental Psychology: General* 133(3), 450-467.

Keller, Frank. 2000. Gradience in grammar: Experimental and computational aspects of degrees of grammaticality. Edinburgh: University of Edinburgh dissertation.

Keller, Frank. 2003. A psychophysical law for linguistic judgments. *Proceedings of the 25th Annual Meeting of the Cognitive Science Society (CogSci–03)*, 652-657.

Lorch, Robert F. & James L. Meyers. 1990. Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology:*

Learning Memory and Cognition 16(1), 149-157.
Luka, Barbara J. & Lawrence W. Barsalou. 2005. Structural facilitation: Mere exposure effects for grammatical acceptability as evidence for syntactic priming in comprehension. *Journal of Memory and Language* 52(3), 436-459.
Noppeney, Uta. & C. J. Price. 2004. An fMRI study of syntactic adaptation. *Journal of Cognitive Neuroscience* 16(4), 702-713.
Pickering, Martin J. & Holly P. Branigan. 1998. The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language* 39(4), 633-651.
Ross, John R. 1967. Constraints on variables in syntax. Cambridge, MA: MIT dissertation. [Published revised as *Infinite Syntax!* [Language and Being], Norwood, NJ: Ablex, 1986.]
Sprouse, Jon. 2007. A program for experimental syntax. College Park, MD: University of Maryland dissertation.
Snyder, William. 2000. An experimental investigation of syntactic satiation effects. *Linguistic Inquiry* 31(3), 575-582.

Jon Sprouse
University of California, Irvine
Department of Cognitive Sciences
3151 Social Science Plaza A
Irvine, CA 92606
USA
jsprouse@uci.edu