A Psychophysical Law for Linguistic Judgments

Frank Keller (keller@inf.ed.ac.uk)
School of Informatics, University of Edinburgh
2 Buccleuch Place, Edinburgh EH8 9LW, UK

Abstract

It has been argued that linguistic acceptability can be estimated using the psychophysical technique of magnitude estimation, in the same way as physical continua such as brightness and loudness (Bard, Robertson, & Sorace, 1996; Cowart, 1997). For physical continua, plotting the perceived stimulus magnitude against the actual physical magnitude results in a power relationship, the Psychophysical Law (Stevens, 1957). We show that a power law of the same kind can be derived by plotting estimated linguistic acceptability against the number of linguistic constraints violated in the stimulus.

Introduction

Magnitude estimation (ME) is a technique standardly applied in psychophysics to measure judgments of sensory stimuli (Stevens, 1957). The procedure requires subjects to estimate the perceived magnitude of a physical stimulus (e.g., the brightness of a light source or the loudness of a beep) by assigning numeric values proportional to stimulus magnitude. Typically, subjects are first presented with a reference stimulus (the modulus), which they assign an arbitrary number. All other stimuli are then judged in proportion to the modulus. For instance, if a stimulus is perceived as twice as bright as the modulus, then the subjects assigns it twice the modulus number, if it is only one third as bright, one third of the modulus number will be assigned.

A simple exponential relationship holds between the physical magnitude of a stimulus (for instance its brightness measured in lux) and its perceived subjective magnitude. Stevens (1957) formulates this as the *Psychophysical Law*:

(1) $\psi = kS^n$

Where ψ is the perceived stimulus magnitude, S is the physical magnitude and k is a constant. Stevens (1957) lists 14 different modalities for which the Psychophysical Law holds; the exponent n is characteristic of a given modality, it can range from .3 for loudness to 2.0 for visual flash rate.[1]

The exponential relationship in (1) can be turned into a linear relationship by log-transforming both ψ and S; this is how psychophysical relationships are typically graphed. An example is given in Figure 1 for brightness and loudness.

---

[1] As an anonymous reviewer points out, Stevens' Law is not uncontroversial. It has been suggested that the power function is an artifact of data averaging across subjects (see Myung, Kim, & Pitt, 2000, and the references cited therein). Also, the power function itself has been claimed to be too versatile: it can fit almost any monotonic curve within the errors of measurement (see Krueger, 1989, and the associated commentaries).
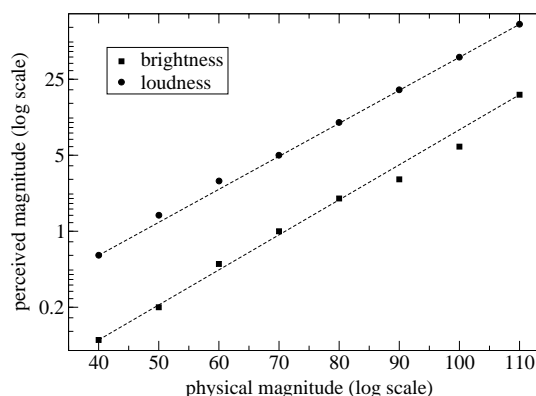

Figure 1: Example of the power law for psychophysical judgment tasks (brightness and loudness of a stimulus); from Stevens (1957)

The ME paradigm has been extended successfully to the psychosocial domain (see Lodge, 1981 for a survey) and recently Bard et al. (1996) and Cowart (1997) showed that linguistic judgments can be elicited in the same way as judgments for sensory or social stimuli. Unlike the five- or seven-point scale conventionally employed in the study of intuitive judgments, ME makes it possible to treat linguistic acceptability as a continuum and directly measures acceptability differences between stimuli. ME has been shown to provide fine-grained measurements of linguistic acceptability, which are robust enough to yield statistically significant results, while being highly replicable both within and across speakers. The techniques has already been applied to wide variety of linguistic phenomena (see Sorace & Keller, 2003 for an overview).

Magnitude estimation of linguistic acceptability is analogous to the standard procedure used to elicit judgments for physical stimuli. Subjects are presented with a series of linguistic stimuli, and have to respond by assigning a numeric value to each stimulus proportional to its perceived acceptability. However, as noted by Bard et al. (1996), the crucial difference between ME of physical stimuli and ME of linguistic stimuli is that for the latter, no objective standard of comparison is available: linguistic acceptability does not have a physical manifestation that can be measured directly.

The aim of the present paper is to address this problem. Our hypothesis is that the theoretical notion of *number of constraint violations* can form the basis of a power law for linguistic judgments analogous to Stevens' Psychophysical Law.

This paper is structured as follows. In the next section, we will review some linguistic background on the phe-

nomenon used as the test case for our power law: word order variation in German. Then we present an ME experiment that elicits data on this phenomenon. We show how the word order data can be accounted for by a power law that relates the number of constraints violated by a sentence to its perceived acceptability. We demonstrate that this law fits the data better than a linear law, and show that this observation extends to a range of results from the literature on ME of linguistic acceptability.

Linguistic Background

To test if the Psychophysical Law can be applied to linguistic judgments, we need a concrete data set. In this paper, we deal with a linguistic phenomenon that has been widely studied in the theoretical literature: word order variation in German. The present section introduces the necessary linguistic background.

Word Order in German

German has a fixed verb order. Subordinate clauses are verb final, while yes/no questions require verb initial order, and declarative main clauses have the verb in second position. In the present experiment we will focus on subordinate clauses, whose order is generally considered the basic one from which the main clause and question orders are derived (e.g., Haider, 1993).[2] An example for the stimuli used in this study is given in (2). We use ditransitive verbs such as *vorschlagen* 'suggest' that can take three animate NPs as complements. For example, the verb final sentence in (2a) is full acceptable, while the verb initial sentence in (2b) is seriously unacceptable.

(2) a. Ich glaube, dass der Produzent dem
       I believe that the producer-NOM the
       Regisseur den Schauspieler vorschlägt.
       director-DAT the actor-ACC   suggests
       'I believe that the producer will suggest the actor to the director.'
    b. Ich glaube, dass vorschlägt der Produzent dem Regisseur den Schauspieler.

While verb order is fixed in German, the order of the complements of the verb is variable. A range of factors can influence the acceptability of the different orders, including case marking, pronominalization, thematic roles, information structure, intonation, definiteness, and animacy (Choi, 1996; Jacobs, 1988; Müller, 1999; Uszkoreit, 1987; Scheepers, 1997). The present study focuses on the effect of case marking and pronominalization on word order, keeping the other factors constant.

We test three different complement orders, with the nominative NP in first, second, and third position, respectively. Examples for these three orders are given in (3). Our notation for word orders uses 'V' for verb, 'S' for subject, and 'O' and 'I' for direct and indirect object, respectively.

(3) a. **SIOV:** Ich glaube, dass der Produzent dem Regisseur den Schauspieler vorschlägt.
    b. **ISOV:** Ich glaube, dass dem Regisseur der Produzent den Schauspieler vorschlägt.
    c. **IOSV:** Ich glaube, dass dem Regisseur den Schauspieler der Produzent vorschlägt.

The experiment is also designed to test the effect of pronominalization on acceptability. The same three orders as in (3) are used, but now one of the NPs is realized as a pronoun. The position of the pronominalized NP varies; either the first, second, or third NP is realized as a pronoun. Example sentences are given in (4) for the order SIOV. We use the index 'pro' to mark the pronominalized NP.

(4) a. $\text{S}_{pro}\text{IOV:}$ Ich glaube, dass er   dem
       I   believe that he-NOM the
       Regisseur den Schauspieler vorschlägt.
       director-DAT the actor-ACC   proposes
       'I believe that he will propose the actor to the director.'
    b. $\text{SI}_{pro}\text{OV:}$ Ich glaube, dass der Produzent ihm den Schauspieler vorschlägt.
    c. $\text{SIO}_{pro}\text{V:}$ Ich glaube, dass der Produzent dem Regisseur ihn vorschlägt.

Word Order Constraints

The fact that different word orders in German differ in their acceptability is typically analyzed in terms of *word order constraints*, i.e., statements on the precedence of constituents that when violated trigger a decrease in acceptability. The effect of constraint violation is typically assumed to be cumulative (Jacobs, 1988; Müller, 1999; Uszkoreit, 1987): the more violations a sentence incurs, the less acceptable it is.

In this paper, we use the set of word order constraints proposed by Uszkoreit (1987, p. 114), listed in (5) (only constraints relevant to the present study are given and constraint names have been added). (Jacobs, 1988; Müller, 1999) use very similar constraints.[3]

(5) a. VERB: $X \prec V[-MC]$
    b. NOM: $[+NOM] \prec [-NOM]$
    c. PRO: $[+PRO] \prec [-PRO]$

Here, '$\prec$' denotes the linear precedence of constituents in a sentence. The constraint VERB relies on the feature MC (main clause) to specify verb order; if this feature is negative (i.e., in a subordinate clause), then the verb has to succeed any other constituent. The constraint NOM requires nominative constituents (marked $[+NOM]$) precede non-nominative constituents (marked $[-NOM]$). The constraint PRO requires pronouns to precede constituents that are not pronouns.

The Experiment

This section reports an experiment was designed to test the hypothesis that the number of constraint violations incurred by a sentence stands in a power law relationship to the perceived acceptability of the sentence. This was

---

[2]Using subordinate clauses avoids potential confounds from topicalization and other phenomena that can occur in verb second clauses. This is standard practice in the psycholinguistic literature on German (e.g., Bader & Meng, 1999).

[3]In fact, the constraint NOM collapses two of Uszkoreit's (1987) constraints and is due to Müller (1999).

tested by eliciting magnitude estimation judgments for sentences that violate between one and five of the word order constraints described in the previous section.

Method

Subjects Thirty-four subjects participated in this experiment, all of them native speakers of German (by self-assessment).

Materials The materials were created based on two subdesigns. The first subdesign included sentences with three full NPs and used a factorial design with the two factors *Nom* and *Verb*, corresponding to violations of the constraints NOM and VERB. Between zero and two violations of NOM were included, and either zero or one violation of VERB. This resulted in an overall design with $Nom \times Verb = 3 \times 2 = 6$ cells. Eight lexicalizations were used per cell. See (2) and (3) for example stimuli.

The second subdesign included sentences with two full NPs and one pronominalized NP. Again, a factorial design was used, this time the additional factor *Pro* was included, corresponding violations of the constraint PRO. Between zero and two violations of PRO were tested. The resulting design had $Nom \times Pro \times Verb = 3 \times 3 \times 2 = 18$ cells. Example stimuli are given in (4). Eight lexicalizations were used for each cell. This resulted in an overall set of 192 stimuli for both subdesigns.

A set of 24 fillers was created, designed to cover the whole acceptability range. A sentence of medium acceptability was used as the modulus item.

Procedure The experimental paradigm was magnitude estimation as described by Stevens (1957) and extended to linguistic stimuli by Bard et al. (1996) and Cowart (1997) (see Introduction for details).

Subjects first saw a set of instructions that explained the concept of numerical magnitude estimation using line length. Subjects were instructed to make length estimates relative to the first line they would see, the reference line. They were told to give the reference line an arbitrary number, and then assign a number to each following line so that it represented how long the line was in proportion to the reference line. Several example lines and corresponding numerical estimates were provided to illustrate the concept of proportionality. Then subjects were told that linguistic acceptability could be judged in the same way as line length. The concept of linguistic acceptability was not defined, but examples of acceptable and unacceptable sentences were provided.

The experiment started with a training phase designed to familiarize subjects with the magnitude estimation task. Subjects had to estimate the length of a set of lines. Then, a set of practice items (similar to the experimental items) were administered to familiarize subjects with applying magnitude estimation to linguistic stimuli. Finally, subjects had to judge the experimental items.

Eight test sets were generated, each containing one lexicalization for each cell in the design, i.e., a total of 24 items. Lexicalizations were assigned to test sets using a Latin square design.

Each subject was randomly assigned one test set. The subject judged 48 items in total: 24 items in the test set and 24 fillers (which were the same for all subjects). Items were presented in random order, with a new randomization being generated for each subject.

Results

The data were normalized by dividing each numerical judgment by the modulus value that the subject had assigned to the reference sentence. This operation creates a common scale for all subjects. All analyses were carried out on the log-transformed normalized judgments, as is standard for magnitude estimation data (Bard et al., 1996; Cowart, 1997).

The aim of this experiment was to test the hypothesis that there is a power relationship between linguistic acceptability and the number of constraint violations. Before we can test this hypothesis, we first have to verify that the factors *Nom*, *Pro*, and *Verb* were effective in implementing the constraints NOM, PRO, and VERB. To this end, we conducted an analysis of variance (ANOVA) for each of the two subexperiments. We will report both by-subject (F_1) and by-item (F_2) analyses.

For the first subexperiment, we found significant main effects of *Nom* ($F_1(2,66) = 28.970$, $p < .0005$; $F_2(2,14) = 19.058$, $p < .0005$) and *Verb* ($F_1(1,33) = 69.816$, $p < .0005$; $F_2(1,7) = 105.594$, $p < .0005$), and a significant interaction of the two factors ($F_1(2,66) = 17.656$, $p < .0005$; $F_2(2,14) = 7.992$, $p = .005$). A post-hoc Tukey test on the factor *Nom* confirmed that zero violations of NOM were more acceptable than a single violation, which in turn was more acceptable than a double violation ($\alpha < .01$ in both cases).

For the second subexperiment, we found significant main effects of *Nom* ($F_1(2,66) = 55.712$, $p < .0005$; $F_2(2,14) = 22.167$, $p < .0005$), *Pro* ($F_1(2,66) = 54.078$, $p < .0005$; $F_2(2,14) = 33.568$, $p < .0005$), and *Verb* ($F_1(1,33) = 63.767$, $p < .0005$; $F_2(1,7) = 851.116$, $p < .0005$). Also all the interactions were significant: *Nom/Pro* ($F_1(4,132) = 7.638$, $p < .0005$; $F_2(4,28) = 4.216$, $p = .009$), *Nom/Verb* ($F_1(2,66) = 30.233$, $p < .0005$; $F_2(2,14) = 14.026$, $p < .0005$), *Pro/Verb* ($F_1(2,66) = 28.871$, $p < .0005$; $F_2(2,14) = 18.705$, $p < .0005$), and *Nom/Pro/Verb* ($F_1(4,132) = 10.907$, $p < .0005$; $F_2(4,28) = 4.105$, $p = .010$). We conducted post-hoc Tukey tests on the factors *Nom* and *Pro*, which confirmed that zero violations were more acceptable than single violations, which in turn were more acceptable than a double violations ($\alpha < .01$ in all cases).

The Psychophysical Law

In this section, we apply Stevens' Psychophysical Law to linguistic judgments, first to the data set obtained in the experiment reported in the previous section, and then to data sets from the literature on magnitude estimation of linguistic acceptability.

Modeling the Experimental Data

In the previous section, we reported significant main effects of the factors *Nom*, *Pro*, and *Verb*, which confirmed that our experimental manipulation was successful in triggering between zero and two violations of the constraints NOM, PRO, and VERB. Each of the sentences in our set of materials violated between zero and three constraints in the first subexperiment (at most two NOM violations and one VERB violation). The second subexperiment included between zero and five constraint viola-
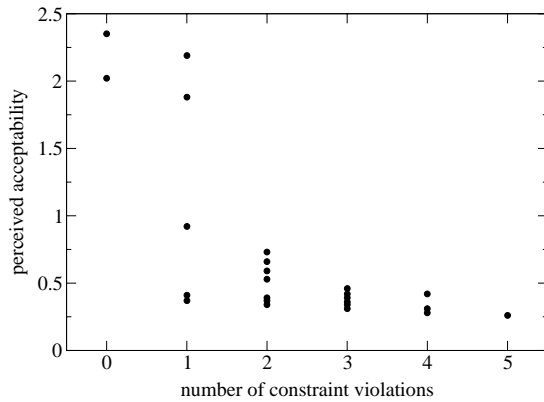
Figure 2: Perceived acceptability as a function of number of word order constraints violated
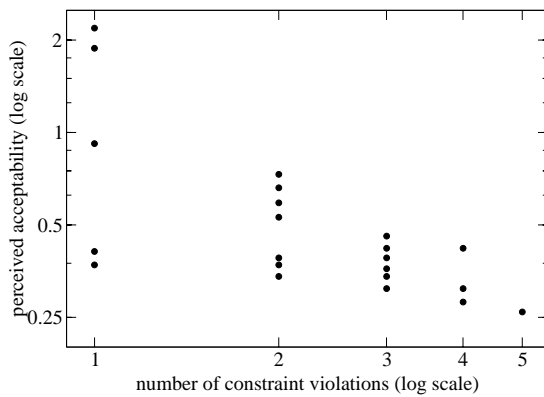

Figure 3: Perceived acceptability as a function of number of word order constraints violated, log-log scale

tions (at most two NOM violations, two PRO violations), and one VERB violation). For further analysis we combined the results of both subexperiment, yielding a set of 24 data points.

The next step was to test the hypothesis that there is a power relationship between the perceived acceptability of a sentence and the number of constraints it violates. Figure 2 plots acceptability against number of violations and clearly shows an exponential relationship (this figure plots the non-log transformed data). As discussed in the introduction, a power relationship like this is characteristic of psychophysical continua such as brightness and loudness. Plotting the data on log-log coordinates should result in a linear relationship. This is illustrated in Figure 3, which plots the same data as Figure 2, now using logarithmic scales (note that the two data points with zero violations were dropped from this graph).

The figures suggest that acceptability behaves like a psychophysical continuum, and that we should be able to fit the psychophysical law in (1) to the data. In contrast to continua like loudness or brightness, linguistic acceptability does not have a natural zero point. Rather, the maximum acceptability of a sentence (corresponding to zero violations) may vary for different linguistic constructions; it may also vary from experiment to experiment, e.g., due to anchoring effects (Nagata, 1992).

To take this into account, we introduced a threshold term l into the equation in (1). This means that we are in effect measuring the exponential *reduction* in acceptability triggered by the constraints a stimulus violates. This yields the following equation:

(6) $\psi = l - kS^n$

Using non-linear regression, we now fitted the terms l, k, and n, which represent the acceptability threshold, the intercept of the exponential function, and its exponent, respectively. The term ψ is the acceptability measured using magnitude estimation, while S is the number of constraint violations. On our data set, non-linear regression yielded a significant relationship between ψ and S ($R = .81$, $N = 24$, $p < .001$), with the following regression equation:

(7) $\psi = 2.21 - 1.21S^{.36}$

Note that the exponent of .36 is within the range of exponents that have been reported for other psychophysical continua, ranging from .3 for loudness to 2.0 for visual flash rate (Stevens, 1957).

As a next step, we tested the hypothesis that acceptability is best described by a power law—one could imagine that a simple linear relationship between the acceptability of a stimulus and the number of constraint violations it incurs fits the data just as well. Indeed, a linear regression on the data yielded a significant predictive relationship ($R = .70$, $N = 24$, $p < .001$). (The linear regression used the equation in (6), but with $n = 1$.)

In order to compare the fit achieved by the two regressions, we computed degrees of freedom adjusted correlation coefficient R'.[4] This adjustment takes into account the number of parameters in the regression equation used to obtain a given R, and therefore makes it possible to compare the fit of the non-linear regression equation (three parameters: l, k, and n) with the fit of the linear regression (only two parameters: l and k). The adjusted correlation coefficients was $R' = .78$ for the non-linear regression and $R' = .67$ for the linear regression. A one-tailed t-test for correlation coefficients showed that the difference between the two R' values was significant ($t(24) = 1.68$, $p < .05$), i.e., the power law yielded a significantly better fit with the data than the linear law.

While the power law in (6) gave rise to a substantial correlation coefficient of $R' = .78$ for our data, there is an obvious problem with the hypothesis that the number of violation directly predicts the acceptability of a sentence: it rests on the assumption that all violations contribute *equally* to acceptability. In our case this means that all three constraints NOM, PRO, and VERB are assumed to cause an equal reduction in acceptability. However, this is not the case, as authors in the theoretical linguistics literature have pointed out (Jacobs, 1988; Müller, 1999; Uszkoreit, 1987). There is also experimental support for the claim that not all the constraints in (5) are equal: Keller (2000a) showed that violations of VERB are more serious than violations of NOM and PRO. Also in our data, the inequality of constraint violations can be observed. Consider the single violations displayed in Figure 2, which represent violations of PRO, NOM, NOM,

---

[4]The formula used was: $R' = \sqrt{1 - (1-R^2)\frac{N-1}{N-k-1}}$, where k is the number of variables in the regression equation.

NOM, and VERB (top to bottom). It is clear that the three constraints differ substantially in the amount of unacceptability they trigger.

To take account of this fact, we elaborated the equation in (6) by introducing *constraint weights* that represent the reduction in acceptability that an individual constraint triggers. This amounts to replacing kS in (6) with a weighted sum, resulting in the following power law:

$$(8) \quad \psi = l - \left(\sum_i w_i C_i \right)^n$$

Here, w_i is the weight of constraint i and C_i is the number of times constraint i is violated. For the data at hand, we obtained the following equation for our three constraints NOM, PRO, and VERB:

$$(9) \quad \psi = l - (w_{\mathrm{Nom}} C_{\mathrm{Nom}} + w_{\mathrm{Pro}} C_{\mathrm{Pro}} + w_{\mathrm{Verb}} C_{\mathrm{Verb}})^n$$

This led to the prediction that equation (9) fits our data better than equation (6), which simply assumes that all constraints have an equal weight (viz., k). We tested this by applying non-linear regression to our data. This resulted in a significant correlation ($R = .89$, $N = 24$, $p < .001$) and the regression equation in (10):

$$(10) \quad \psi = 2.25 - (1.38 C_{\mathrm{Nom}} + .84_{\mathrm{Pro}} C_{\mathrm{Pro}} + 3.78 C_{\mathrm{Verb}})^{.38}$$

Note that the number of parameters differs for the two non-linear equations: the weighted equation in (9) contains five parameters, while the unweighted equation in (6) only contains three. In order to compare the fit achieved by these two equations, we again computed degrees of freedom adjusted correlation coefficients: $R' = .78$ and $R' = .85$ for (7) and (10), respectively. The difference between the two adjusted coefficients was significant ($t(24) = 1.64$, $p < .05$), which means that a power law based on weighted constraints achieved a better fit on the data than a power law with unweighted constraints.

Figures 4 and 5 plot acceptability scores against the weighted sum of the number of constraint violations, on linear and logarithmic scales, respectively. The fact that the data points cluster around a straight line if they are plotted on a log-log scale confirms the hypothesis that perceived acceptability is subject to a power law, just like other psychophysical continua.

Finally, we carried out a linear regression analysis using equation (9), but setting $n = 1$. This resulted in a significant correlation ($R = .78$, $N = 24$, $p < .001$). The adjusted correlation coefficient of $R' = .73$ was significantly lower than the $R' = .85$ obtained for the weighted power law ($t(24) = 2.62, p < .01$). This demonstrates that a power law yielded a better fit than a linear law, even for weighted constraints.

Generalizing to other Data Sets

As mentioned in the Introduction, each modality is characterized by a specific exponent n in the psychophysical power law. The hypothesis that linguistic acceptability is a psychophysical continuum like loudness and brightness makes an important prediction: there should be a unique exponent for acceptability that is invariant across experiments. This hypothesis can be tested by applying the power law in (8) to data sets from the literature. Table 1 presents the results of regression analyses using equation (8) for seven data sets from the literature; all
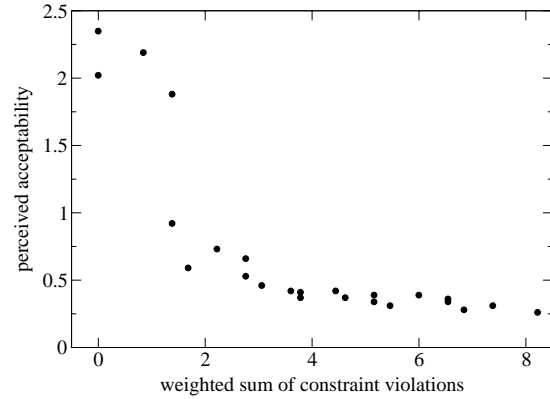

Figure 4: Perceived acceptability as a function of the weighted sum of the number of word order constraints violated
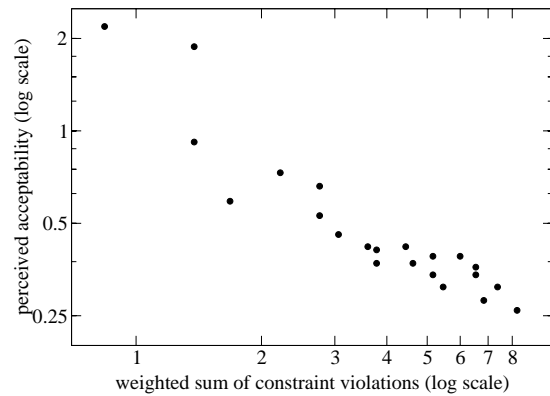

Figure 5: Perceived acceptability as a function of the weighted sum of the number of word order constraints violated. log-log scale

studies are magnitude estimation studies of linguistic acceptability. We only included data from experiments that were based on an explicit set of linguistic constraints, as this is the prerequisite for applying the power law. Note that some of the experiments included context as a between-subject variable. We conducted separate analyses for the context and the no context condition, as the context has an influence on the threshold l for acceptability (for example, Keller & Alexopoulou, 2001 found that judgments are higher in the no context condition, all other factors being equal).

These published data cover a range of syntactic constructions (word order, extraction, gapping) in three different languages (German, Greek, English). The results show that the weighted power law provides a consistently good fit with the experimental data. The fit of the power law is significantly better than the fit of the corresponding linear law in all but two cases (see Table 1). The exponent of the power law, n, ranges from .14 to .83, the average is $n = .36$. Note that there are two outliers: the data of Experiment 1 (context) and Experiment 2 (no context) of Keller and Alexopoulou (2001) results in exponents of $n = .14$ and $n = .83$ respectively. These are also the two cases where the fit of the power law is not significantly

Table 1: Applying the power law to data sets in the literature

Construction	Language	N	m	R'_l	R'_p	p	n	Source
Word order	German	16	3	.79	.92	**	.22	Keller, 2000a, Exp. 1, no context
Word order	German	24	4	.84	.90	*	.34	Keller, 2000a, Exp. 1, context
Word order	Greek	24	3	.72	.81	–	.14	Keller & Alexopoulou, 2001, Exp. 1, context
Word order	Greek	24	3	.88	.87	–	.83	Keller & Alexopoulou, 2001, Exp. 2, no context
Word order	Greek	36	5	.81	.87	*	.30	Keller & Alexopoulou, 2001, Exp. 2, context
Extraction	English	16	6	.80	.93	**	.28	Keller, 2000b, Exp. 4
Gapping	English	24	3	.83	.89	*	.43	Keller, 2001, Exp. 2, context

N	number of data points		m	number of constraints
R'_l	adj. correlation coefficient for linear law		p	sig. difference R'_l and R'_p ($*p < .05$; $**p < .01$)
R'_p	adj. correlation coefficient for power law		n	exponent for power law

better than that of the linear law. In the case of $n = .83$ this is expected, as the exponent is close to one, resulting in basically a linear law. If we discount these two outliers, then the remaining exponents are all close to the value of $n = .38$ that we found when applying the power law to the experimental data set reported in this paper.

Conclusions

This paper dealt with the question of whether linguistic acceptability can be treated as a psychophysical continuum such as brightness and loudness. It is well known that linguistic acceptability can be measured using the psychophysical technique of magnitude estimation. However, the psychophysical power law that relates perceived magnitude to physical magnitude is not directly applicable, as linguistic acceptability has no physical correlate.

We therefore proposed a power law that relates the perceived acceptability of a linguistic structure to the number linguistic constraints that the structure violates. We presented experimental data for word order variation in German that allowed us to test this hypothesis. It was found that a power law closely models the experimental data; a comparison with a linear law relating acceptability and number of violations yielded a significantly worse fit. We were also able to show that a modified power law that assigns weights to constraints yields an even better fit with the experimental data. Again, it significantly outperforms the corresponding linear law.

Finally, we generalized our results by applying them to a range of data sets from the literature on magnitude estimation of linguistic acceptability. In all cases a close fit with the data was achieved, and in all but two cases the power law provided a better fit than the linear law. We also found that the modality-specific exponent in the Stevens' Psychophysical Law is approximately $n = .36$ for linguistic acceptability.

References

Bader, M., & Meng, M. (1999). Subject-object ambiguities in German embedded clauses: An across-the-board comparison. *Journal of Psycholinguistic Research*, *28*(2), 121–143.

Bard, E. G., Robertson, D., & Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language*, *72*(1), 32–68.

Choi, H.-W. (1996). *Optimizing structure in context: Scrambling and information structure.* Unpublished doctoral dissertation, Stanford University.

Cowart, W. (1997). *Experimental syntax: Applying objective methods to sentence judgments.* Thousand Oaks, CA: Sage Publications.

Haider, H. (1993). *Deutsche Syntax generativ: Vorstudien zur Theorie einer projektiven Grammatik.* Tübingen: Gunter Narr.

Jacobs, J. (1988). Probleme der freien Wortstellung im Deutschen. In I. Rosengren (Ed.), *Sprache und Pragmatik* (Vol. 5, pp. 8–37). Department of German, Lund University.

Keller, F. (2000a). Evaluating competition-based models of word order. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 747–752). Mahwah, NJ: Lawrence Erlbaum Associates.

Keller, F. (2000b). *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality.* Unpublished doctoral dissertation, University of Edinburgh.

Keller, F. (2001). Experimental evidence for constraint competition in gapping constructions. In G. Müller & W. Sternefeld (Eds.), *Competition in syntax* (pp. 211–248). Berlin: Mouton de Gruyter.

Keller, F., & Alexopoulou, T. (2001). Phonology competes with syntax: Experimental evidence for the interaction of word order and accent placement in the realization of information structure. *Cognition*, *79*(3), 301–372.

Krueger, L. E. (1989). Reconciling Fechner and Stevens: Toward a unified psychophysical law. *Behavioral and Brain Sciences*, *12*(2), 251–267.

Lodge, M. (1981). *Magnitude scaling: Quantitative measurement of opinions.* Beverley Hills, CA: Sage Publications.

Müller, G. (1999). Optimality, markedness, and word order in German. *Linguistics*, *37*(5), 777–818.

Myung, I. J., Kim, C., & Pitt, M. (2000). Toward an explanation of the power law artifact: Insights from response surface analysis. *Memory and Cognition*, *28*(5), 832–840.

Nagata, H. (1992). Anchoring effects in judging grammaticality of sentences. *Perceptual and Motor Skills*, *75*(1), 159–164.

Scheepers, C. (1997). *Menschliche Satzverarbeitung: Syntaktische und thematische Aspekte der Wortstellung im Deutschen.* Unpublished doctoral dissertation, University of Freiburg.

Sorace, A., & Keller, F. (2003). Gradience in linguistic data. *Lingua*, to appear.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, *64*(3), 153–181.

Uszkoreit, H. (1987). *Word order and constituent structure in German.* Stanford, CA: CSLI Publications.