BUCLD 44 Proceedings
To be published in 2020 by Cascadilla Press
Rights forms signed by all authors

Testing the Tolerance Principle on Corpus Data

Xiaomeng Ma, Qihui Xu, Virginia Valian and Martin Chodorow

1 Introduction

1.1 Deriving the Tolerance Principle

Do children learn a rule for forming the past tense? Or do they simply have a list of verbs and local regularities? If they do have a rule, how is it formed, given that the input is noisy? Yang (2016) has proposed the Tolerance Principle (TP) to explain how children deploy general (productive) rules given noisy input. We first describe the principle and then our revision of it. Yang assumes that humans apply the Elsewhere Condition (e.g. McClelland and Rumelhart, 1981; Plaut, 1997) in processing rules and exceptions. The Elsewhere Condition can be implemented as a serial search procedure in which each lexical item is first compared to all the exceptions to the general rule. If a match is found, a specific rule for the matching exception is triggered. If not, the general rule is applied, as shown in (1):

(1) IF w = e_1 THEN apply rule e_1 to w...
IF w = e_2 THEN apply rule e_2 to w...
IF w = e_3 THEN apply rule e_3 to w...
...
IF w = e_N THEN apply rule e_N to w...
IF w not in set{e_1,e_2,e_3,e_4....e_N}, THEN apply general rule to w...

For example, to retrieve the past tense form for the verb *eat*, *eat* is first compared to the exceptions in the irregular verb inventory. When it is found, the irregular form for *eat* is applied to derive the past tense form *ate*. To retrieve the past tense form for the verb *type*, when the search for a match in the irregular verb set fails, the general past tense rule is applied to derive *typed*.

Since lexical retrieval is a self-terminating serial search process, rules and exceptions are organized to minimize the time required. A productive rule to produce a set of N items is derived when applying the rule takes less time than does processing all the items individually. Without a productive rule, all the items are ranked by their frequencies, so that frequently used items will be processed most quickly. When a productive rule is generated, items are separated into two categories, regulars and exceptions. The exceptions are ranked by frequency and the rule is

applied only when the item is not found in the set of exceptions. For example, when there is no productive rule and the vocabulary inventory has n regular items (w) and m irregular items (e), all the items are arranged according to their frequency, as shown in (2). When there is a productive rule, all the irregular items are ranked based on frequency, and all the regular items are concatenated into a set where one rule will be applied, as shown in (3).

(2) Without productive rule: (3) With productive rule:

$$N=n+m\begin{cases} w_1 & frequency:100 \\ e_1 & frequency:99 \\ w_2 & frequency:98 \\ \ldots & frequency:\ldots \\ w_m & frequency:2 \\ e_n & frequency:1 \end{cases} \qquad N=n+1\begin{cases} e_1 & frequency:99 \\ e_2 & frequency:90 \\ e_3 & frequency:87 \\ \ldots & frequency:\ldots \\ e_n & frequency:1 \\ & w_1,w_2,w_3\ldots w_n \end{cases}$$

The rule will be generated only when the time to process the list with the rule is less than processing all the items. As shown in (2) and (3), there are fewer items (n+1) to process in the list with a productive rule than in the list without a productive rule (n+m). On occasion, however, the time to search for an item in the list *with* a rule can be more than the search time in the list *without* a rule. For example, for the item w_2, which is a frequent regular word, it will take less time to find w_2 in the list without a productive rule than in the list with one. Since w_2 is the third most frequent word, the search only needs to examine w_1 and e_1 before finding a match for w_2. In the list with a productive rule, the regular item will be reached only after all the exceptions are checked, which creates more time to find w_2. Therefore, the rule will be deployed only when the average time complexity to search for each word in the list with a productive rule (T_R) is less than the list without a rule (T_N).

(4) $T_N > T_R$

According to Yang (2016), the average time to search for a word is the product of the probability of the word and the retrieval time for the word. To calculate the probability for each word, Yang assumes that any sample from a large corpus follows the Zipfian distribution; therefore, the product of the frequency of the word (f_i) and the rank (r_i) of the word is a constant (C).

(5) $r_i f_i = C_i$

The probability of a word in a corpus (p_i) is the frequency of the word (f_i) divided by the sum of the frequencies of all the words. The probability of occurrence (p_i) for w_i can be expressed as (6) where $H_N = \sum_{k=1}^{N} \frac{1}{r_k}$.

(6) $$p_i = \frac{f_i}{\sum_{k=1}^{N} f_k} = \frac{\frac{C_i}{r_i}}{\sum_{k=1}^{N} \frac{C_k}{r_k}} = \frac{\frac{1}{r_i}}{\sum_{k=1}^{N} \frac{1}{r_k}} = \frac{1}{r_i \cdot H_N}$$

Since all the items are stored in a list ranked according to frequency, the retrieval time for each item is determined by the rank of the item. For example, the word *the* appears more frequently in the corpus than the word *give*, and therefore it takes requires less time to retrieve. Yang simplifies this as 'the i-th ranked item takes i units of time to be retrieved' (Yang, 2018b, p.???). The time for w_i (T_{w_i}) is shown in (7). The total time for N items in the list (T_N) is shown in (8).

$$(7) \quad T_{w_i} = r_i \cdot \frac{1}{r_i \cdot H_N} = \frac{1}{H_N}$$

$$(8) \quad T_N = \sum_{k=i}^{N} (r_i \cdot \frac{1}{r_i \cdot H_N}) = \frac{N}{H_N}$$

If a rule is used, the exceptions are stored in a ranked list and the regular items are stored in a set after the list of exceptions. The exception list is processed in the same way as is the list without rules. If there are e items in the exceptions, the total time(T_{e_i}) for the exception list is shown in (9). The total time for all the items in the exception list (T_e) is shown in (10).

$$(9) \quad T_{e_i} = r_i \cdot \frac{1}{r_i \cdot H_e} \cdot \frac{e}{N} = \frac{e}{N \cdot H_e}$$

$$(10) \quad T_e = r_i \cdot \frac{1}{r_i \cdot H_e} \cdot \frac{e}{N} \cdot e = \frac{e \cdot e}{N \cdot H_e}$$

The time for all regulars is the same, which is the constant e, given that there are e items in the exception list. The total time to process all the regular items (T_w) is:

–

$$(11) \quad T_{\overline{w}} = (1 - \frac{e}{N}) \cdot e$$

The total time for a list with a productive rule (T_R) is the sum of the time for exceptions and regular items, as shown in (12):

$$(12) \quad T_R = T_e + T_{\overline{w}} = \frac{e}{N} \cdot \frac{e}{H_e} + (1 - \frac{e}{N}) \cdot e$$

A productive rule will be derived only when T_R is smaller than T_N. To derive the maximum number for the exceptions (e), first we approximate the Nth harmonic number with the natural log ($\ln N$), then we make $T_R \leq T_N$:

$$
\begin{aligned}
& T_R \leq T_N \\
& \frac{e}{N} \cdot \frac{e}{H_e} + (1 - \frac{e}{N}) \cdot e \leq \frac{N}{H_N} \\
(13) \quad & \frac{e}{N} \cdot \frac{e}{lne} + (1 - \frac{e}{N}) \cdot e \leq \frac{N}{lnN} \\
& \frac{e^2}{N} \cdot (\frac{1}{lne} - 1) + e \leq \frac{N}{lnN}
\end{aligned}
$$

Since $\frac{e^2}{N} \cdot (\frac{1}{lne} - 1)$ is always smaller than or equal to zero, as long as e is smaller than $\frac{N}{lnN}$, then T_R is always smaller than T_N. Thus, the TP is derived:

(14) *Tolerance Principle*
Let R be a rule applicable to N items, of which e are exceptions. R is productive if and only iff:
$e \leq \theta_N, where \theta_N = \frac{N}{lnN}$ (Yang, 2016, p.64)

1.2 Why should(n't) the Tolerance Principle (TP) work?

Recall that the TP assumes the Elsewhere Condition: lexical access time is roughly logarithmic (Murray and Forster, 2004). But there is no guarantee that it is executed as a cognitive function of learning. The TP appealingly handles a critical fact about language acquisition: children generate rules based on noisy input. The TP provides an elegant and succinct way to quantify the 'noise' in the input.

Evidence from artificial language learning (Schuler et al., 2016) supports the TP. Children between the ages of 5 and 7 heard names for nine novel objects in both singular and plural forms. Each plural marker either followed a rule (add *ka*) or instead used an individual suffix (add *po, tay, lee bae, muy,* or *woo*). In one condition, children heard five nouns with the *ka* marker and four with individual markers. In another condition, they heard three nouns with the *ka* marker and six with individual markers. As the TP predicts, children learned under the 5/4 condition but not the 3/6 condition, as shown by their ability to use *ka* as a general plural marker in a Wug-like test.

Despite this evidence, other research has queried whether the TP can be applied to explain language acquisition in real life.

Yang used $\ln N$ as the approximation of the Harmonic number, as shown in (13). For smaller N, however, $\log_2 N$ is a better approximation of HN (Andreae & Kuiper, 2018). For example, in Schuler

et al. (2016)'s experiment, when N = 9, $H_9 \approx 2.83$, while $\ln 9 \approx 2.20$ and $\log_2 9 \approx 2.71$. For N = 9, the base-2 logarithm is a better approximation than the natural logarithm. That the values matter can be seen by looking at the 4 exceptions, which were calculated based on $\theta = N/lnN$, where 9/2.20 $\approx$ 4.10. If, instead, the original $H_9 \approx 2.83$ were inserted, then $\theta = N/H_N$ would be 9/2.83 $\approx$ 3.18. That result produces a tolerance threshold of three exceptions, which goes against the results in Schuler et al.

The TP also received criticism on the more general level of how to approach acquisition. For example, perhaps communicative needs, context, prior learning and cognitive load should be taken into account (Goldberg, 2018). As another example, serial search might be a flawed model compared to a more relevant lexical retrieval model (Kapatsinski, 2018).

In this paper we use corpus data to test the TP and we explore a method to use corpus data.

2 Testing the Tolerance Principle on Corpus Data

2.1 Yang's Test on Adam's and Eve's Data

In Chapter 4 of Yang (2016)'s book, he established a procedure for the applying TP. In our study, we also follow his procedure.

(15) a. Obtain a rule *R* along with its structural description and structural change.

b. Count *N*, the number of lexical items that meet the structural description of *R*.

c. Count *e*, the subset of *N* that are exceptions to R.

d. Compare *e* and the critical threshold $\theta_N = N/lnN$ to determine productivity.

Yang first applied this procedure in explaining the acquisition of past tense in English. English speaking children usually start to produce the past-tense form by the age of 2. Most children also produce overregularization errors on past tense, such as *grewed*, *feeled* (e.g. Marcus et al., 1992). The first instance of an overregularization error can be seen as an unambiguous marker for the presence of a productive 'add *-d*' rule for past tense.

Adam produced his first overregularization error at the age of 2;11, when he said *What dat feeled like?* (Brown, 1973). This error implied that Adam had already constructed the past tense rule. According to TP, Adam must have had a

[1] See Yang (2018a) for his defense.

large enough corpus of verbs (N) that the irregular verbs (e) could be tolerated, and e was smaller than $\theta = N/lnN$. Adam's first recording starts at 2;3. Yang thus estimated Adam's effective vocabulary (N) as all the verbs he produced between 2;3 and 2;11. Yang did not only count the past tense of the verbs, he counted all forms of verbs. According to Yang, as long as Adam produced one form of a verb, that verb has to be in Adam's lexicon. Based on this method, he found 300 verbs, which made N = 300. Therefore, $\theta = N/lnN \approx 53$, which means children can learn the rule when there are fewer than 53 irregular verbs out of 300 verbs. However, Yang counted 57 irregular verbs in Adam's total 300 verb lexicon. He attributed the difference between 57 and 53 to sampling effects.

Yang also used the same method to test Eve's data. Eve's first overregularization error appeared at 1;10 when she said *it falled in the briefcase*[2] (Brown, 1973). Yang found 163 verbs Eve produced between 1;6, when Eve had her first recording, and 1;10. When $N = 163$, $\theta = N/lnN \approx 32$, which means Eve could only tolerate 32 irregular verbs when using the past tense rule. However, Yang found 49 irregulars in her production, which is again higher than the prediction. He attributed the difference to undersampling of Eve's data.

2.2 Revised Testing Methodology

In Yang's test, TP failed to account for Adam's and Eve's data on past tense acquisition. With the proposed new methodology, we aim to preserve Yang's insight but develop a different version of TP.

First, we seek to develop a better estimate of the effective vocabulary (N) of the child. In Yang's method, he used the raw count of the verbs children produced as the N. There are two problems with this estimation. First, children's production only reflects part of their effective vocabulary as children do not utter all the words they know. Second, the estimated effective vocabulary should not be an exact number; instead, it would be more accurate to use a possible range for N. We propose to estimate N through parents' input (U_p) and children's production (U_c), both of which can be extracted from the corpora. Since children do not absorb parents' input completely, and since their productions do not represent their entire linguistic knowledge, we introduce λ to represent comprehension cost (%) and δ to represent production loss (%). λ and δ should range between 0% - 100%, which can also be roughly estimated from the corpora by calculating the overlap rate of parents' vocabulary and children's vocabulary. In addition, in order to compensate for the loss of the data due to undersampling, we introduce X_c and X_p for the words that are not picked up in recording sessions for the child and the parent, respectively. The estimated N can be written as below:

$$N = (U_p + X_p) \cdot \lambda = \frac{(U_c + X_c)}{\delta} \quad (16)$$

[2] Yang might have made an error in his book. Eve made the first overregularization error at the age of 1;8, when she uttered *I seed it* meaning *I saw it*.

Second, we aim to produce a more accurate measure ranked frequency distribution of N. Yang assumed a Zipfian distribution for all items N and all the exceptions e. However, a Zipfian distribution is not guaranteed for a small corpus, such as all the verbs a 2-year-old child knows. Not having a perfect Zipfian distribution will make formula (5) invalid, thus affecting all derivations following it. Recall formula (5): when a corpus follows a Zipfian distribution, the product of the frequency of the word (f_i) and the rank (r_i) is a constant C, shown in (17) again. This is derived from the formal expression of Zifp's law, which is shown in (18): the frequency of the rth most frequent word is inversely proportional to its rank. The Zifpian distribution is a special case of a power law function where the exponent (α) equals 1. Formula (5) is only valid when α is 1, which fits the Zifpian distribution. When a smaller corpus (such as children's effective vocabulary of verbs and irregular verbs) does not have a Zifpian distribution and the exponent (α) is not 1, formula (5) is no longer valid, thus all the following derivations will not hold true.

(17) $r_i \cdot f_i = C$ (replicate of (5))

(18) $f_i = C r_i^{-\alpha}, \alpha = 1$

In this paper, we propose to measure the actual corpus distributions of all the verbs (N) and the irregular verbs (e), and use in our calculations the empirically estimated exponents from the equations that best fit those ranked frequency distributions. Since the distributions of all the verbs and the irregular verbs are not necessarily the same, we will use α and β for the exponents, respectively. To include the exponent as a variable, formulas (6) to (13) can be rewritten as follows:

(19) Probability of occurrence for ith ranked word (p_i):

$$p_i = \frac{\frac{1}{r_i^{\alpha}}}{\sum_{k=1}^{N}(\frac{1}{r_k^{\alpha}})} = \frac{1}{r_i^{\alpha} \cdot H_{N,\alpha}}$$

(20) Time complexity for a list of N items without a productive rule (T_N):

$$T_N = \sum_{k=1}^{N}(r_i \cdot \frac{1}{r_i^{\alpha} \cdot H_{N,\alpha}}) = \frac{H_{N,\alpha-1}}{H_{N,\alpha}}$$

(21) Time complexity for a list with e exceptions and a productive rule (T_R):

$$T_R = \frac{H_{e,\beta-1}}{H_{e,\beta}} \cdot \frac{e}{N} + (1 - \frac{e}{N}) \cdot e$$

(22) A productive rule will be derived when $T_R \leq T_N$:

$$\frac{H_{e,\beta-1}}{H_{e,\beta}} \cdot \frac{e}{N} + (1 - \frac{e}{N}) \cdot e \quad \leq \frac{H_{N,\alpha-1}}{H_{N,\alpha}}$$

Unlike formula (13) where H_N can be conveniently approximated using lnN, there is no mathematical approximation for the Harmonic number in the inequation (22). Therefore, inequation (22) can not be used to estimate the value of θ,

the maximum number of irregular verbs children can tolerate before they learn the past tense rule. Instead of comparing the estimated number of irregular verbs and observed irregular verbs in children's vocabulary, we propose to compare T_R and T_N directly. In the next section, we extract all the variables (e, N, α and β) from three children's corpora to compare T_N and T_R. The Tolerance Principle will be confirmed if T_R is smaller than T_N as predicted in (22).

2.3 Testing on Adam's, Eve's and Fraser's Data

In this section, we use the revised testing methods to test Adam's, Eve's (Brown, 1973) and Fraser's data (Lieven et al., 2009) on their past tense acquisition. As mentioned above, Adam's recording starts at the age of 2;3, and he made the first overregularization error at the age of 2;11. Eve's recording starts at the age of 1;6, and she made the first overregularizaiton error at the age of 1;8. Fraser's recording starts at 2;0, and he made the first overregularization error at the age of 2;5 (Maslen et al., 2004). A summary of their corpus data is shown in table 1.

Table 1: Summary of Adam's, Eve's and Fraser's Data

	Adam	Eve	Fraser
Age of first recording	2;3	1;6	2;0
Age of first overregularization error	2;11	1;8	2;5
No. of files in between	18	5	90

All of the data were automatically extracted from the annotated corpora in CHILDES using the NLTK python package. The verbs in each file were identified using part-of-speech taggers annotated by the MOR program (MacWhinney, 2012). The numbers of verbs and irregular verbs in parents' input (U_p and e_p) and in children's production (U_c and e_c) are shown in table 2. The distributions of all the verbs and the irregular verbs were fitted onto power law functions. The exponents for all the verbs in parents' input (α_p), in children's production (α_c) and the exponents for all the irregular verbs in parents' input (β_p) and children's production (β_c) are shown in table 3. The log-log graphs can be found in the Appendix.

To compare the time complexity for the list with a productive rule (T_R) and the list without a productive rule (T_N), e, α, β and U were inserted to the inequation in (22). In this test, we simplified the process of estimating the effective vocabulary (N) by using the verbs produced by children (U_c) and their parents (U_p) as the Ns. If the time complexity for a list with a rule (T_R) is smaller than for a list without a rule (T_N), the Tolerance Principle will be confirmed. The number of the time complexity for each child is shown in table 4 and 5.

Adam's and Fraser's data support the TP's prediction, but Eve's data do not, as shown in table 4 and 5. This could be attributed to the undersampling effect,

Table 2: Counts of verbs and irregular verbs

	Adam	Eve	Fraser
U_p	275	136	566
U_c	270	91	358
e_p	70	50	97
e_c	62	36	78

Table 3: Exponents for the distributions of verbs and irregular verbs

	Adam	Eve	Fraser
α_p	0.69	0.74	0.56
α_c	0.66	0.84	0.60
β_p	0.64	0.65	0.44
β_c	0.61	0.73	0.49

Table 4: Comparing T_R vs T_N using U_c as N

Production	T_R	T_N	$T_R \leq T_N$
Adam	52.5	78.1	True
Eve	26.3	22.6	**False**
Fraser	67.3	110.6	True

Table 5: Comparing T_R vs T_N using U_p as N

Input	T_R	T_N	$T_R \leq T_N$
Adam	57.9	76.2	True
Eve	37.7	36.9	**False**
Fraser	86.7	181.6	True

since Eve only had 5 files between the first recording and the first appearance of an overregularization error, while Adam had 18 files and Fraser had 90 files. In order to further test Eve's data, we set $T_R = T_N$ in order to estimate a minimum number for N, and compared the observed number of verbs U to N. TP will be confirmed if U and N are reasonably close. Formula (23) is used to calculated the minimum N by inserting e, α, β for Eve's production and Eve's mother's input. When N is estimated using Eve's production, $f(120) \sim 0$, which means that Eve has to produce at least 120 verbs in order to acquire the rule. She was observed to have produced 91 verbs in 5 files, with an average of 18 verbs per file. To observe at least 120 verbs, we only need to less than two files. Thus the minimum N predicted by TP is close enough to the observed U for Eve's production. When Eve's mother's input was used to estimate N, $f(141) \sim 0$. Eve's mother has already been observed to have produced 136 verbs, which is close enough to 141. The result is shown in table 6. The estimated minimum N is close enough to the observed U, thus TP can be confirmed on Eve's data too.

(23) Minimum N to confirm TP when $T_R = T_N$:

$$f(N) = T_R - T_N = \frac{H_{e,\beta-1}}{H_{e,\beta}} \cdot \frac{e}{N} + (1 - \frac{e}{N}) \cdot e - \frac{H_{N,\alpha-1}}{H_{N,\alpha}} = 0$$

3 Conclusion

Yang's Tolerance Principle is built on two fundamental assumptions: 1) lexical retrieval is a serial search process, 2) the distribution of children's effective vocabulary is Zipfian. This paper tested the second assumption on children's ef-

Table 6: Comparing the minimum N and the observed U

	Eve's production	Eve's mother's input
U	91	136
N	120	141
$U \sim N$	Yes	Yes

fective vocabulary of verbs and irregular verbs and revised the methodology so that it is testable on corpus data. Adam's, Eve's and Fraser's data have shown that the distribution of children's effective vocabulary follows a general power law distribution, but not Zipf's law. Therefore, we revised Yang's formula by introducing the exponent of the power law as a variable. We then tested it on three children's data (Adam, Eve and Fraser). The results supported TP's prediction.

Appendix

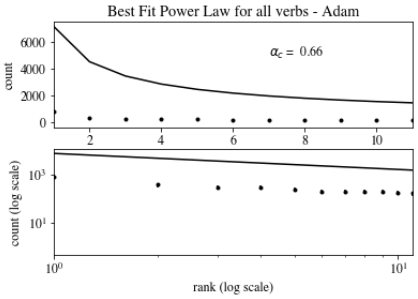

Figure 1: Distribution of Adam's Verbs

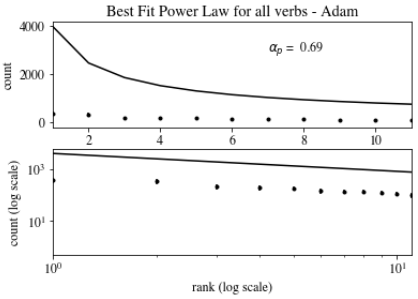

Figure 2: Distribution of Adam's mother's verbs

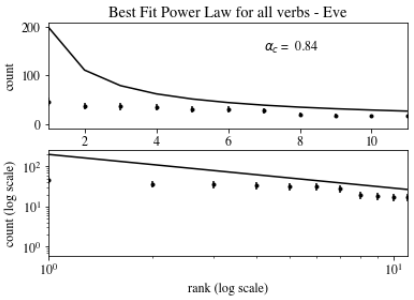

Figure 3: Distribution of Eve's Verbs

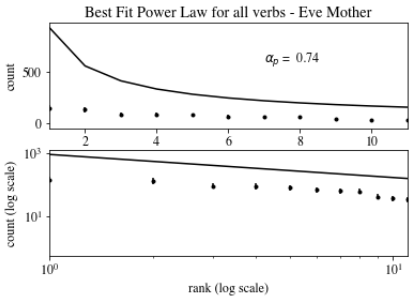

Figure 4: Distribution of Eve's mother's verbs

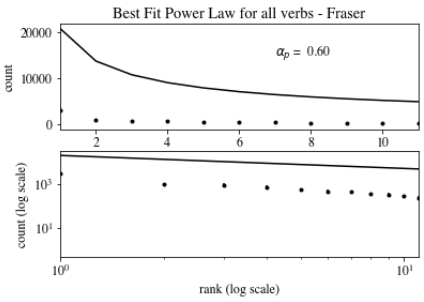

Figure 5: Distribution of Fraser's Verbs

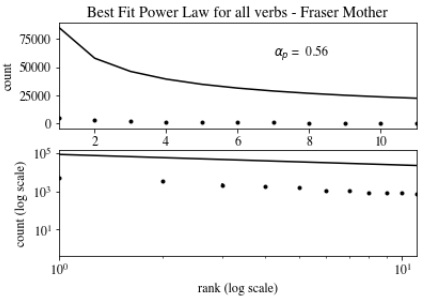

Figure 6: Distribution of Fraser's mother's verbs

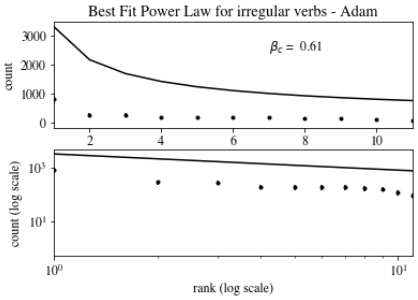

Figure 7: Distribution of Adam's Irregular Verbs

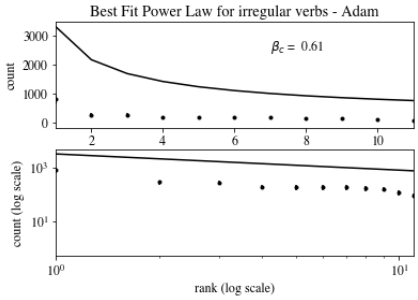

Figure 8: Distribution of Ada's mother's Irregular verbs

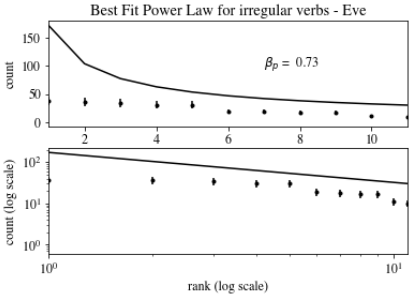

Figure 9: Distribution of Eve's Irregular Verbs

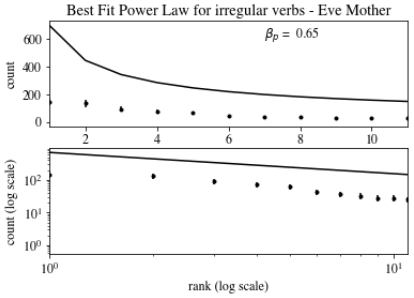

Figure 10: Distribution of Eve's mother's Irregular verbs

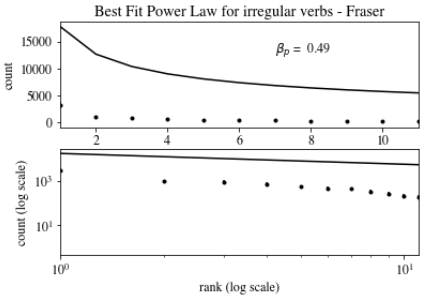

Figure 11: Distribution of Fraser's Irregular Verbs

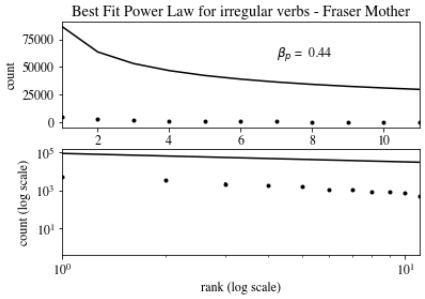

Figure 12: Distribution of Fraser's mother's Irregular verbs

References

John H Andreae and Koenraad Kuiper. Charles yang's tolerance principle. 2018.

Roger Brown. 1973: A first language: the early stages. cambridge, ma: Harvard university press. 1973.

Adele E Goldberg. The sufficiency principle hyperinflates the price of productivity. *Linguistic Approaches to Bilingualism*, 8(6):727–732, 2018.

Vsevolod Kapatsinski. On the intolerance of the tolerance principle. *Linguistic Approaches to Bilingualism*, 8(6):738–742, 2018.

Elena Lieven, Dorothé Salomo, and Michael Tomasello. Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics*, 20(3):481–507, 2009.

Brian MacWhinney. Morphosyntactic analysis of the childes and talkbank corpora. In *LREC*, pages 2375–2380, 2012.

Gary F Marcus, Steven Pinker, Michael Ullman, Michelle Hollander, T John Rosen, Fei Xu, and Harald Clahsen. Overregularization in language acquisition. *Monographs of the society for research in child development*, pages i–178, 1992.

Robert JC Maslen, Anna L Theakston, Elena VM Lieven, and Michael Tomasello. A dense corpus study of past tense and plural overregularization in english. *Journal of Speech, Language, and Hearing Research*, 2004.

James L McClelland and David E Rumelhart. An interactive activation model of context effects in letter perception: I. an account of basic findings. *Psychological review*, 88(5):375, 1981.

Wayne S Murray and Kenneth I Forster. Serial mechanisms in lexical access: the rank hypothesis. *Psychological Review*, 111(3):721, 2004.

David C Plaut. Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and cognitive processes*, 12(5-6):765–806, 1997.

Kathryn D Schuler, Charles Yang, and Elissa L Newport. Testing the tolerance principle: Children form productive rules when it is more computationally efficient to do so. In *CogSci*, 2016.

Charles Yang. *The price of linguistic productivity: How children learn to break the rules of language*. MIT Press, 2016.

Charles Yang. Some consequences of the tolerance principle. *Linguistic Approaches to Bilingualism*, 8(6):797–809, 2018a.

Charles Yang. A user's guide to the tolerance principle. *Manuscript. University of Pennsylvania (ling. auf. net/lingbuzz/004146)*, 2018b.