Regularization or lexical probability-matching?
How German speakers generalize plural morphology

Kate McCurdy **Sharon Goldwater** **Adam Lopez**

School of Informatics
University of Edinburgh

Abstract

Artificial language learning research has shown that, under some conditions, adult speakers tend to *probability-match* to inconsistent variation in their input, while in others, they *regularize* by reducing that variation. We demonstrate that this framework can characterize speaker behavior in a natural-language morphological inflection task: the lexicon can be used to estimate variation in speaker productions. In the task of German plural inflection, we find that speakers probability-match a lexical distribution conditioned on phonology, and largely disregard an alternative possible strategy of conditional regularization based on grammatical gender.

Keywords: Regularization; Morphology; Natural language

Introduction

Research in artificial language learning shows that adult speakers have a range of responses to unpredictable inconsistencies in their linguistic input. Under some circumstances, they *probability-match* and reproduce the variation in their input distribution, while in other circumstances they prefer to *regularize*[1] and produce more consistent patterns (Hudson Kam, 2019). Often, speakers regularize by increasing production of the most frequent variant in their input (Hudson Kam & Newport, 2005, 2009). For example, given the hypothetical artificial language training vocabulary and test inputs shown in Table 1, Speaker A shows this kind of regularization: the *-s* plural form has gone from 60% of their input to 100% of their output, thus reducing variation in how the plural form is realized. By contrast, Speaker B matches the input probability of *-s*, keeping it at 60% in their output. The learning biases influencing speaker behavior in these experiments are not fully understood, and show complex interactions with communicative pressures in cultural transmission (Smith et al., 2017). As a result, it is challenging to anticipate which artificial language findings will apply in more complex natural language environments, such as the German plural inflection task we explore in this study.

Research on natural language variation shows that it is typically *conditioned* upon multiple factors, both linguistic (e.g. phonological environment) and non-linguistic (e.g. speaker identity) (Chambers & Schilling, 2018). Conditional variation provides another mechanism for regularization: unpredictable variation can become predictable when conditioned on particular linguistic contexts. In Table 1, Speaker C shows this type of regularization, consistently mapping "le" articles to *-s* and "ze" articles to null forms.

Singular Article + Noun	Plural
le dug	dugs
le gat	gats
le brid	brids
ze pik	piks
ze cheep	cheep
ze bish	bish

Test Input	Speaker A	Speaker B	Speaker C
le gee	gees	gees	gees
ze koo	koos	koos	koo
ze teer	teers	teer	teer

Table 1: Hypothetical artificial language. Top table shows training vocabulary, bottom shows test outputs from three speakers. A regularizes overall variation, B probability-matches, and C regularizes conditional variation.

Note that our example artificial language experiment frames variation with respect to static attributes within a lexicon: each individual noun has two fixed classes (expressed by the article and the plural form), and we consider how speakers might use membership in one class (e.g. article) as a cue to signal membership in another class (plural form). Artificial language learning studies have shown that adult speakers can learn to condition noun class assignment on such markers when they are statistically reliable (Frigo & McDonald, 1998). Culbertson, Gagliardi, and Smith (2017) found that learners may prefer different cues to noun class (e.g. phonological vs. semantic cues) based on salience or early availability in training. While they found reliable statistical main effects from their experimental cue manipulation, their data show a broad range of variation within individuals as well, suggesting the type of variation in speaker strategies illustrated by the hypothetical case in Table 1.

The studies discussed above explore speaker generalization using toy lexicons, where the amount and type of variation can be manipulated experimentally. However, in principle it should be possible to apply some of the same analysis methods to the more complex case of generalization from natural language. In particular, German number inflection pro-

[1] N.B. we use the term *regularize* in the linguistics sense (reduce variation), not the machine learning sense (reduce overfitting).

vides a complex natural-language test case for the type of lexical variation seen in our hypothetical experiment. Some aspects of the German plural system are well-described by rules (e.g. derived nouns; Augst, 1979), but other parts of the lexicon show more complex probabilistic relations, and psycholinguistic experiments reveal considerable variation between speakers when they are asked to produce the plural forms of novel words (Mugdan, 1977; Köpcke, 1988; Zaretsky & Lange, 2016; McCurdy, Goldwater, & Lopez, 2020).

In this work, we adopt the framework of probability-matching versus regularization to shed light on this variability. We ask whether variation in German number inflection of novel words can be explained in terms of a) lexical statistics and b) variation in individual speaker strategies. Do speakers predominantly probability-match to the distribution observed in the lexicon, leading to the variation observed in behavioral experiments? Or do they predominantly regularize, but with different speakers pursuing different strategies (e.g. reducing conditional vs. overall variation) which lead to a general appearance of inconsistent behavior?

We use the information-theoretic definition of regularization presented by Ferdinand, Kirby, and Smith (2019) to evaluate individual behavior in terms of entropy. We take the joint distribution of grammatical gender (G) and plural inflection class (C) observed in the lexicon as a reference distribution to assess German speaker behavior on a dual task: for each of 24 novel nouns, identify its grammatical gender, and produce its plural inflected form. We find that, consistent with some artificial language experiments, adult speakers largely probability-match the conditional variation observed in the input, and disregard an alternative strategy of gender-conditioned regularization. Our work shows that lexical statistics across items can predict speaker behavior within novel items, connecting artificial language findings with natural language behavior.

Background

German number inflection Here we present a highly simplified overview of the German plural system, illustrated with reference to the CELEX2 lexical dataset (Baayen, Piepenbrock, & Gulikers, 1995). Each German noun has two lexical attributes relevant to our analysis: its grammatical gender (G) and plural inflection class (C). A noun can have masculine (M), neuter (N), or feminine (F) gender, and this lexical property has a complex relation to the noun's phonology and semantics (Köpcke & Zubin, 1984). Gender is indicated on the article which precedes the noun it its singular form.[2]

The other key lexical attribute, plural inflection class, is indicated by the plural form of the noun. This is typically characterized by at least five predominant suffixes, which can be combined with umlaut[3] to give eight classes (Mugdan,

[2]Following much of the literature, we consider only nouns in their citation form, i.e. only in nominative case.

[3]Umlaut is a process by which back vowels are fronted; for example, the noun *Apfel* "apple" takes the plural form *Äpfel*. Following a smaller subset of the literature, we ignore umlaut in this analysis.

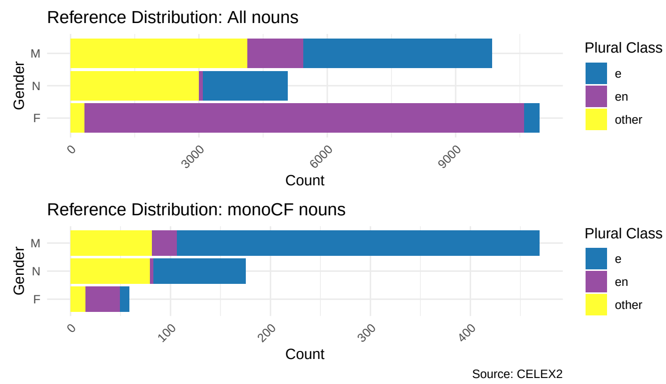

Figure 1: Reference distributions calculated from German CELEX2. Top, all nouns (excluding ambiguous gender nouns). Bottom, monosyllabic consonant-final nouns.

1977), although more fine-grained distinctions are possible — CELEX2 labels 13 separate plural inflection classes. Nonetheless, two suffixes predominate: 45% of noun types in the CELEX2 lexicon take the *-(e)n* suffix in the plural, and 26% take *-e*. As regularization often involves increasing frequent variants, we focus on these two suffixes and their relationship to grammatical gender. Fig. 1a shows the joint distribution of 3 simplified plural class (*-e*, *-(e)n*, and "other") by gender over **all nouns** in CELEX2. Fig. 1b focuses on the subset of nouns in CELEX2 with a similar phonological shape to our experimental stimuli, i.e. monosyllabic and consonant-final (**monoCF**).

Gender and plural class Our key research question is whether German speakers will regularize overall variation, probability match the observed lexical distribution, or regularize conditional variation. In the latter case, grammatical gender is the most viable option on which to condition plural class, for several reasons. 1) There is a clear strong statistical relationship between gender and plural class, evident in Fig. 1. Williams et al. (2020) analyze a subset of German nouns in CELEX2, and estimate that 25% of the variation in inflection class (including all plurals and cases) can be explained by grammatical gender. For our simplified set of inflection classes, we estimate 40% (Table 2). 2) Many linguists have analyzed grammatical gender as the primary determinant of plural class, with *-e* as the default class for non-feminine nouns, and *-(e)n* for feminine nouns (e.g. Augst, 1979; Wiese, 1999; Bittner, 1999). 3) Neural models of German inflection reliably learn to condition plural class on gender (Goebel & Indefrey, 2000; McCurdy, Lopez, & Goldwater, 2020; Dankers, Langedijk, McCurdy, Williams, & Hupkes, 2021). Despite this, many psycholinguistic studies report little (Köpcke, 1988; Zaretsky & Lange, 2016; McCurdy, Lopez, & Goldwater, 2020) or no (Mugdan, 1977; Marcus, Brinkmann, Clahsen, Wiese, & Pinker, 1995; Spreng, 2004) effect of gender on speaker productions.

Regularization and German plurals Our entropy-based framing of regularization leads us to focus on the two most frequent inflection classes. This contrasts with some of the literature on German plural inflection, where different theories of regularization have emphasized the role of minority classes. For example, Marcus et al. (1995) argue that only the rare plural suffix *-s* is regular in the sense that it is rule-generated, while under their analysis other plural classes are not. This Dual Route interpretation has been highly influential (see e.g. Clahsen, 1999, and replies), but its claims have been challenged empirically (e.g. Zaretsky & Lange, 2016; Behrens, 2017; McCurdy, Goldwater, & Lopez, 2020). Other analyses of the German plural system have focused on productivity and type frequency, either with a rule-based analysis (e.g. Yang, 2016) or without (e.g. Köpcke, 1988; Bybee, 1995; Heitmeier, Chuang, & Baayen, 2021).

Herce (2019) notes that the term "regularity" is associated with many distinct concepts in the linguistics literature, and recommends that researchers use more precise language, e.g. "productivity" or "predictability." Our approach emphasizes the "predictability" dimension, in line with other recent attempts to formalize an information-theoretic concept of morphological regularity (Ackerman & Malouf, 2013; Cotterell, Kirov, Hulden, & Eisner, 2018; Wu, Cotterell, & O'Donnell, 2019). Note, however, that these analyses use the lexicon to estimate the *regularity* of a lexical item, for example to predict that the English past tense form "jumped" is more regular (i.e. predictable) than "ran." In contrast, we use the lexicon to assess *regularization* behavior by speakers: do they maintain the level of variation present in the lexicon, or introduce more predictability to novel lexical items?

Methods

Quantifying regularization

Ferdinand et al. (2019) present a novel quantitative analysis of regularization in terms of entropy. Under their definition, speaker regularizing behavior is formalized as the degree of entropy reduction relative to a reference distribution. All measures here originate with Shannon (1948).

The first key measure is Shannon entropy, which quantifies in bits the complexity, or variation, over the distribution of a single categorical variant. In our case, we're interested in entropy over plural class C:

$$H(C) = -\sum_{c \in C} \mathrm{P}(c) \log_2 \mathrm{P}(c) \tag{1}$$

Similarly, we calculate $H(G)$ to obtain the entropy of the distribution over grammatical gender.

The second key measure is conditional entropy, which calculates the entropy of our variant of interest C conditioned on grammatical gender G:

$$H(C \mid G) = -\sum_{g \in G} \mathrm{P}(g) \sum_{c \in C} \mathrm{P}(c \mid g) \log_2 \mathrm{P}(c \mid g) \tag{2}$$

	$H(G)$	$H(C)$	$MI(C;G)$	$U(C \mid G)$
All nouns	1.52	1.54	0.61	40%
All (6 cl.)	1.52	1.98	0.67	34%
monoCF	1.19	1.21	0.18	14%
mCF (6 cl.)	1.19	1.55	0.23	15%

Table 2: CELEX2 entropy measurements for gender $H(G)$, plural class $H(C)$, mutual information between plural class and gender $MI(C;G)$, and percentage plural variation explained by gender $U(C|G)$. We see similar values whether using our simplified 3-class analysis or a more traditional 6-class analysis for C.

Subtracting conditional entropy from Shannon entropy gives the mutual information between the two variables:

$$MI(C;G) = MI(G;C) = H(C) - H(C \mid G) \tag{3}$$

The mutual information can be normalized by the Shannon entropy to get an estimate of the percentage of variation explained by the conditioning variable, known as the uncertainty coefficient (Williams et al., 2020)[4]:

$$U(C \mid G) = \frac{MI(C;G)}{H(C)} = \frac{H(C) - H(C \mid G)}{H(C)} \tag{4}$$

Under Ferdinand et al.'s framework, any reduction in entropy relative to the reference distribution qualifies as regularization. They note that this can be accomplished in three ways: reducing variation in either the distribution of the variant $H(C)$, or of the context $H(G)$, or the conditional distribution $H(C|G)$ (equivalent to increasing $MI(C;G)$).

Reference distribution In artificial language learning experiments, the reference distribution is typically defined by the researcher and manipulated as an experimental variable. By contrast, in the current study, we use the entropy metrics defined above to compare speaker behavior to two references: the distribution of grammatical gender and plural inflection class over a) **all nouns** in the German lexicon, and b) **monoCF nouns**, i.e. monosyllabic nouns ending in a consonant, as these are phonologically similar to our stimuli. We use the CELEX German lexicon (Baayen et al., 1995) to calculate these reference distributions, shown in Fig. 1.

The measures $H(C)$, $H(G)$, $MI(C;G)$, and $U(C \mid G)$ for the reference distribution are reported in Table 2. By default, we use the simplified 3-class categorization (*-e*, *-(e)n*, "other") for inflection C. We additionally report entropy measures with a more traditional 6-class categorization (*-e*, *-(e)n*, *-er*, *-∅*, *-s*, and "other") to show that including minority classes doesn't substantially alter the analysis.

[4]We thank an anonymous reviewer for noting that mutual information is typically normalized with respect to the smaller entropy, in this case $H(G)$. We use $H(C)$ in the denominator as we are specifically interested in $U(C \mid G)$, the fraction of plural class entropy explained by gender, rather than the inverse relation $U(G \mid C)$.

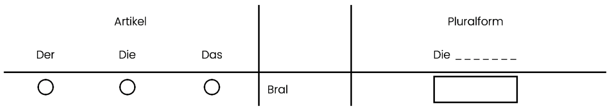

Figure 2: Task presentation for one item. To the left of the novel noun is gender selection, to the right, written plural.

Behavioral experiment

Stimuli The stimuli used in this experiment comprise 24 monosyllabic nouns ending in a consonant (i.e. monoCF nouns), originally developed by Marcus et al. (1995). As seen in Fig. 1 and Tab. 2, this class of nouns is ambiguous in terms of plural class and grammatical gender. This makes them good candidates to assess regularizing behavior — other phonological classes of German nouns already have fully predictable inflection class assignments, e.g. nouns ending in schwa near-universally take the *-(e)n* plural. These stimuli have also been used in multiple previous experiments (Marcus et al., 1995; Zaretsky & Lange, 2016), so our results can be straightforwardly compared with previous findings.

Task The task is a version of the well-known wug test (Berko, 1958): participants were given a novel noun, such as *wug* (or in our case the more Germanic *Vag*), and asked to produce its plural inflected form. Our experiment includes an additional dimension. Along with the plural form, participants were asked to indicate the presumed grammatical gender of the noun by selecting the corresponding article for its singular form, as shown in Fig. 2.

We had two motivations for adding the gender task. Firstly, as earlier wug test studies have found weak to absent effects of gender on German plural inflection (Mugdan, 1977; Marcus et al., 1995; Spreng, 2004; Zaretsky & Lange, 2016; McCurdy, Lopez, & Goldwater, 2020), we sought an experimental design which would compel participants to attend to the gender of the noun. Secondly, we wanted participants to generate the full joint distribution over grammatical gender (G) and inflection class (C), so that we could evaluate their regularization behavior with respect to all three strategies identified by Ferdinand et al.

Procedure After providing consent, participants completed an onboarding task, in which they had to provide the gender and plural form for 12 real German nouns. Participants had to answer these questions correctly to proceed to the experiment. After the onboarding, participants were randomly assigned to one of three lists counterbalanced for presentation order of gender (e.g. “Der/Die/Das Vag” v.s. “Das/Der/Die Vag”). Within each list, the 24 test items were presented in randomized order. We publicly release the data.[5]

[5] https://github.com/kmccurdy/german-wug-data/

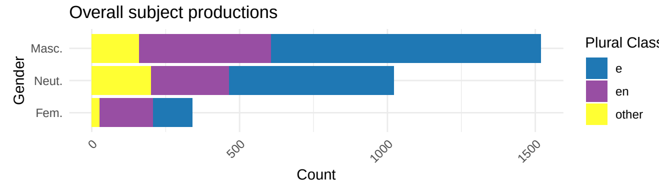

Figure 3: Gender and plural productions from participants. Compare to reference distributions in Fig. 1.

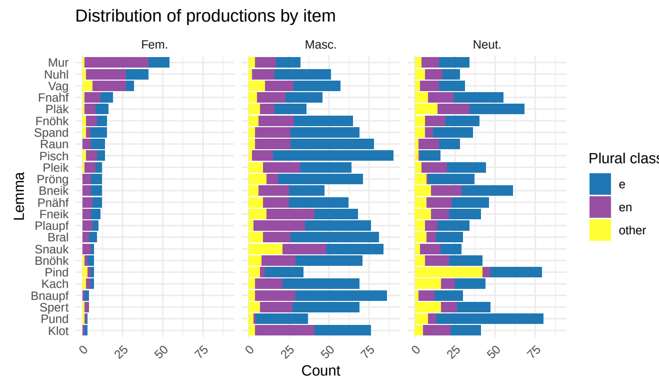

Figure 4: Gender and plural productions by item.

Participants We recruited 120 speakers with German as a first language to complete an online survey using the platform Prolific.[6] Speakers were compensated at the rate-adjusted equivalent of 11.50 USD per hour.

Analysis Following Ferdinand et al. (2019), we quantify the entropy in the distribution produced per participant, and use it to classify participant behavior. Ferdinand et al. assume that participants with entropy measures within the 95% confidence interval (CI) bounds show behavior consistent with *probability-matching* the relevant distribution. To define probability-matching behavior, we simulate experimental draws over 24 items by sampling from the relevant joint categorical distributions. For each reference distribution, we first sample 10^5 grammatical gender assignments for the items, then plural class assignments conditional on the sampled gender. We calculate a more conservative 90% CI by taking the 5th and 95th percentiles of the resulting simulations. Participants with entropy measures below the 5% CI bound are classified as *regularizers*, and above the 95% are *variabilizers*, with respect to the same distribution.

We build on Ferdinand et al.'s approach by also considering the *type* of regularization observed: overall reduction in variation (i.e. reducing $H(C)$) versus conditional reduction in variation (reducing $H(C \mid G)$, i.e. increasing $MI(C;G)$ or its normalized equivalent $U(C \mid G)$).

[6] https://www.prolific.co

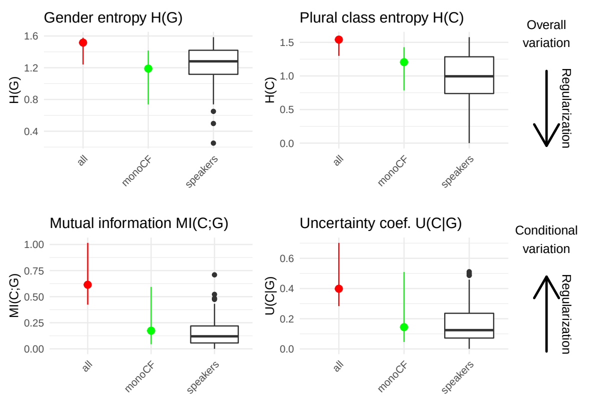

Figure 5: 90% CIs for reference distributions (all and monoCF nouns), and observed values for speakers. Speakers may regularize overall variation in $H(C)$, but do not appear to regularize $MI(C;G)$.

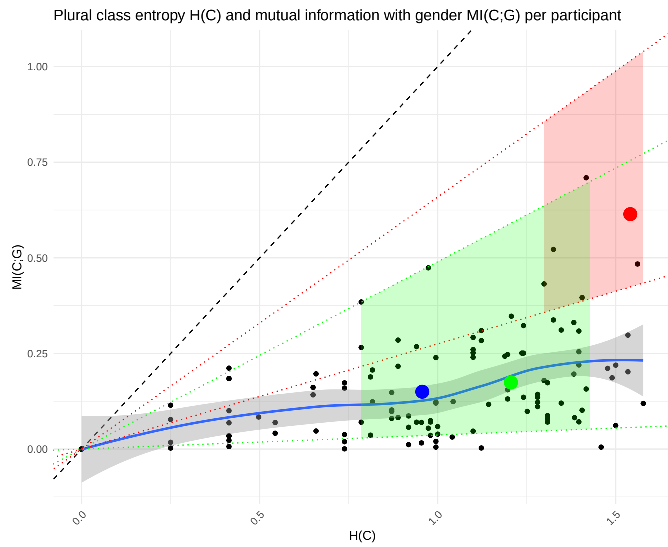

Figure 6: H(C) and MI(C;G) by participant. Color boxes and lines show 90% CI for all (red) and monoCF (green) nouns; color dots show reference values. The blue dot shows the speaker grand mean, and the blue line shows a Loess regression fit of speaker MI(C;G) on H(C). Most participants are in the green box, consistent with probability-matching the monoCF noun distribution.

Results and Discussion

Fig. 3 presents the overall distribution of gender and plural productions from all participants (compare to the reference distributions in Fig. 1). We see considerable variation in gender and plural class assignment, which does not appear to be driven by strong item-level biases (c.f. Fig. 4).

Do speakers regularize overall variation? Fig. 5 shows the 90% CIs for the two reference distributions, and the observed range of speaker values, for our entropy-based measures. Overall variation is shown in the top row. For gender $H(G)$, most speakers' productions are consistent with probability-matching either reference distribution, falling within both CIs. For plural class $H(C)$, we see some evidence for regularization: 75% of speakers reduce variation below the all-nouns 5% CI bound (c.f. Tab. 3). The bulk of those speakers show variation consistent with probability-matching the monoCF distribution, although 27% also fall below the 5% CI bound. In sum, we have two possible interpretations: either speakers are insensitive to the phonological properties of the stimuli and a large majority regularize plural class (i.e., relative to the lexicon as a whole); or speakers condition on phonology and are mainly probability-matching to a phonologically similar subset of the lexicon. However, the further analysis below suggests that speakers *are* sensitive to phonology, which makes the latter interpretation more plausible.

Do speakers regularize conditional variation? The lower row of Fig. 5 shows 90% CIs and the observed distribution for the conditional variation measures $MI(C;G)$ and $U(C \mid G)$, where higher values indicate greater predictability given the conditioning factor. Here we have clear evidence that speakers do *not* regularize by conditioning on grammatical gender; in fact, they seem to be probability-matching to the level of gender-conditioned predictability found in the monoCF nouns, which is substantially lower than that of the full lexicon. Speakers could, in principle, use the stronger relationship between gender and inflection class found in the full lexicon to make predictions about the stimuli, but they do not. This result is surprising given the importance of gender in both linguistic analyses (e.g. Augst, 1979; Wiese, 1996; Bittner, 1999) and recent models (e.g., recent neural network models make predictions that are consistent with the level of gender conditioning in the full lexicon; Goebel & Indefrey, 2000; McCurdy, Lopez, & Goldwater, 2020; Dankers et al., 2021). Our information-theoretic analysis suggests that speakers in fact condition on phonological form at the expense of predictability due to gender.

Interestingly, this reduced level of gender conditioning $MI(C;G)$ appears consistent relative to plural variation $H(C)$, although it need not be: speakers who vary plural class more could in principle introduce more gender conditioning. Fig. 6 shows, for each individual participant, how much variation they produced over plural class $H(C)$ — farther right on the x-axis indicates a more varied set of plural classes — and how much that variation was influenced by grammatical gender $MI(C;G)$ — higher on the y-axis indicates more gender-conditioning, i.e. a tighter statistical coupling between gender and plural class. The dotted black line shows $MI(C;G) = H(C)$, the theoretical maximum statistical coupling: a point on that line would represent a speaker whose plural class assignments were fully explained by grammatical gender, for example always assigning masculine nouns to the

All nouns	Var. $H(C)$	Prob.-match	Regl.
Variabilize $U(C\|G)$	1% (1)	21% (25)	**60% (72)**
Probability-match		3% (4)	10% (12)
Regularize			
N/A ($H(C)=0$)			5% (6)

monoCF nouns	Var. $H(C)$	Prob.-match	Regl.
Variabilize $U(C\|G)$	2% (2)	8% (10)	3% (4)
Probability-match	6% (7)	**58% (69)**	18% (21)
Regularize			1% (1)
N/A ($H(C)=0$)			5% (6)

Table 3: Speaker strategy classification. Relative to all nouns, most speakers regularize overall plural class variation while *increasing* variability with respect to gender (upper table). Relative to monoCF nouns, most speakers probability-match overall and gender-conditioned plural class variation (lower).

-e plural class and feminine nouns to the *-(e)n* class. We see that even speakers who produce as much plural class variation as observed in the lexicon ($H(C) > 1.3$) are mostly below the red box, meaning their gender-conditioning $MI(C;G)$ is more typical of the monoCF distribution.

General Discussion

Our findings demonstrate that the regularization/probability-matching framework developed in the artificial language learning literature can also describe behavior in natural language tasks. Our work is not the first to show this; Hendricks, Miller, and Jackson (2018) used this framework to study variable gender assignment in a Germanic dialect, finding that some children regularized while others probability-matched the variation in the adult distribution. To the best of our knowledge, however, we are the first to use lexical statistics as a reference distribution to evaluate regularization behavior in a psycholinguistic experiment.

We suspect that probability-matching lexical statistics provides a stronger account for our results than most formal models. The substantial variation within items (c.f. Fig. 4) suggests a fundamental incompatibility with any models that make strong item-level predictions, which would encompass most rule-based models (e.g. Mugdan, 1977; Yang, 2016). Exemplar-based models (e.g. Hahn & Nakisa, 2000) may better handle such variability, but doing so appears to require extensive fine-tuning (c.f. Rosen, 2022). As noted earlier, parts of the German plural system are readily described by rules — our findings apply to the subset of the lexicon which shows less predictability. That said, many linguistic accounts of German inflection have proposed high-level rules based on grammatical gender (e.g. August, 1979; Wiese, 1996; Bittner, 1999), and neural models of German inflection learn behavior consistent with such rules (Goebel & Indefrey, 2000; McCurdy, Lopez, & Goldwater, 2020; Dankers et al., 2021). Our findings challenge such accounts: speakers *could* regularize by conditioning on gender to the extent observed in the German lexicon as a whole (i.e. 40% of plural class variation, c.f. Tab. 2), but instead they match the lower level of gender conditioning typical of the phonological class (15-16%). This accords with other linguistic accounts which consider gender subordinate to phonology (e.g. Mugdan, 1977; Spreng, 2004). Furthermore, our study's experimental design explicitly foregrounds gender by forcing participants to select both the article and plural class for each noun. This means that our results likely represent a *ceiling* for gender conditioning on these stimuli. Previous studies with the same stimuli have presented the article instead, and found weaker or absent effects of gender (Marcus et al., 1995; Zaretsky & Lange, 2016; McCurdy, Lopez, & Goldwater, 2020).

Conditional variation seems to play a paradoxical role in these results. On the one hand, there is strong evidence for phonological conditioning: speaker behavior is consistent with the lexical statistics of a phonologically similar subset of the lexicon, rather than the lexicon as a whole. On the other hand, we have two mysteries. Firstly, this phonological conditioning only appears at the level of word class; phonology does not seem to drive strong biases for individual items. Secondly, this phonological conditioning comes at the expense of gender conditioning: participants make grammatical gender *less* informative than it is in the lexicon as a whole. It is unclear how these trends relate to artificial language learning studies, which have found that adult learners tend to condition on lexical identity (i.e. reducing variation across nouns by assigning each noun to one lexical class; Smith & Wonnacott, 2010; Samara, Smith, Brown, & Wonnacott, 2017). Johnson, Culbertson, Rabagliati, and Smith (2020) find that high mutual information (i.e. low i-complexity; Ackerman & Malouf, 2013) benefits learning for neural networks, but not for speakers, while low overall entropy (i.e. low e-complexity) benefits both. Our results echo their findings, as speakers appear to reduce overall entropy ($H(C)$), but unlike neural models, do not increase mutual information ($MI(C;G)$).

Conclusion

In this work, we take an information-theoretic measure of regularization developed for artificial language learning research, and use it to analyze experimental results in the natural-language domain of German plural inflection. We consider two possible points of reference — the lexicon of German nouns as a whole, and a restricted subset with a particular phonological shape — and find that speaker behavior is best described as probability-matching the lexical statistics of the latter phonologically-conditioned distribution. Although speakers could plausibly regularize by conditioning on grammatical gender (as predicted by the statistics of the overall lexicon), instead they appear to probability-match the lower level of gender conditioning seen on phonologically similar nouns. We demonstrate that lexical statistics can predict how speakers generalize lexical attributes to novel items, connecting artificial language findings with natural language behavior.

References

Ackerman, F., & Malouf, R. (2013). Morphological Organization: The Low Conditional Entropy Conjecture. *Language*, *89*(3), 429–464. (Publisher: Linguistic Society of America) doi: 10.1353/lan.2013.0054

Augst, G. (1979). Neuere Forschungen zur Substantivflexion. *Zeitschrift für Germanistische Linguistik*, *7*, 220. (Place: Berlin Publisher: Walter de Gruyter.)

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX2 lexical database. *Linguistic Data Consortium*.

Behrens, H. (2017). The role of analogy in language processing and acquisition. *The changing English language: Psycholinguistic perspectives*, 215–239. (Publisher: Cambridge University Press Cambridge)

Berko, J. (1958, August). The Child's Learning of English Morphology. *WORD*, *14*(2-3), 150–177. doi: 10.1080/00437956.1958.11659661

Bittner, D. (1999). Gender classification and the inflectional system of German nouns. In *Gender in Grammar and Cognition: I: Approaches to Gender. II: Manifestations of Gender* (pp. 1–24). Berlin ; New York: De Gruyter Mouton.

Bybee, J. (1995, October). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*(5), 425–455. doi: 10.1080/01690969508407111

Chambers, J. K., & Schilling, N. (2018). *The handbook of language variation and change*. John Wiley & Sons.

Clahsen, H. (1999, December). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences*, *22*(6), 991–1013. doi: 10.1017/S0140525X99002228

Cotterell, R., Kirov, C., Hulden, M., & Eisner, J. (2018). On the complexity and typology of inflectional morphological systems. *Transactions of the Association for Computational Linguistics*, *7*, 327–342.

Culbertson, J., Gagliardi, A., & Smith, K. (2017, February). Competition between phonological and semantic cues in noun class learning. *Journal of Memory and Language*, *92*, 343–358. doi: 10.1016/j.jml.2016.08.001

Dankers, V., Langedijk, A., McCurdy, K., Williams, A., & Hupkes, D. (2021, November). Generalising to German Plural Noun Classes, from the Perspective of a Recurrent Neural Network. In *Proceedings of the 25th Conference on Computational Natural Language Learning* (pp. 94–108). Online: Association for Computational Linguistics.

Ferdinand, V., Kirby, S., & Smith, K. (2019, March). The cognitive roots of regularization in language. *Cognition*, *184*, 53–68. doi: 10.1016/j.cognition.2018.12.002

Frigo, L., & McDonald, J. L. (1998, August). Properties of Phonological Markers That Affect the Acquisition of Gender-Like Subclasses. *Journal of Memory and Language*, *39*(2), 218–245. doi: 10.1006/jmla.1998.2569

Goebel, R., & Indefrey, P. (2000). A recurrent network with short-term memory capacity learning the German-s plural. *Models of language acquisition: Inductive and deductive approaches*, 177–200.

Hahn, U., & Nakisa, R. C. (2000, December). German Inflection: Single Route or Dual Route? *Cognitive Psychology*, *41*(4), 313–360. doi: 10.1006/cogp.2000.0737

Heitmeier, M., Chuang, Y.-Y., & Baayen, R. H. (2021, June). Modeling morphology with Linear Discriminative Learning: considerations and design choices. *arXiv:2106.07936 [cs]*. (arXiv: 2106.07936)

Hendricks, A. E., Miller, K., & Jackson, C. N. (2018, January). Regularizing Unpredictable Variation: Evidence from a Natural Language Setting. *Language Learning and Development*, *14*(1), 42–60. doi: 10.1080/15475441.2017.1340842

Herce, B. (2019). Deconstructing (ir)regularity. *Studies in Language*, *43*(1), 44–91. doi: 10.1075/sl.17042.her

Hudson Kam, C. L. (2019, October). Reconsidering Retrieval Effects on Adult Regularization of Inconsistent Variation in Language. *Language Learning and Development*, *15*(4), 317–337. doi: 10.1080/15475441.2019.1634575

Hudson Kam, C. L., & Newport, E. L. (2005, April). Regularizing Unpredictable Variation: The Roles of Adult and Child Learners in Language Formation and Change. *Language Learning and Development*, *1*(2), 151–195. doi: 10.1080/15475441.2005.9684215

Hudson Kam, C. L., & Newport, E. L. (2009, August). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, *59*(1), 30–66. doi: 10.1016/j.cogpsych.2009.01.001

Johnson, T., Culbertson, J., Rabagliati, H., & Smith, K. (2020, March). *Assessing Integrative Complexity as a predictor of morphological learning using neural networks and artificial language learning* (preprint). PsyArXiv. doi: 10.31234/osf.io/yngw9

Köpcke, K.-M. (1988, April). Schemas in German plural formation. *Lingua*, *74*(4), 303–335. doi: 10.1016/0024-3841(88)90064-2

Köpcke, K.-M., & Zubin, D. A. (1984). *Sechs Prinzipien für die Genuszuweisung im Deutschen: ein Beitrag zur natürlichen Klassifikation*.

Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R., & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive psychology*, *29*(3), 189–256.

McCurdy, K., Goldwater, S., & Lopez, A. (2020, July). Inflecting When There's No Majority: Limitations of Encoder-Decoder Neural Networks as Cognitive Models for German Plurals. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 1745–1756). Online: Association for Computational Linguistics.

McCurdy, K., Lopez, A., & Goldwater, S. (2020, November). Conditioning, but on Which Distribution? Grammatical Gender in German Plural Inflection. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics* (pp. 59–65). Online: Association for Computational Linguistics. doi: 10.18653/v1/2020.cmcl-1.8

Mugdan, J. (1977). *Flexionsmorphologie und Psycholinguis-*

tik: Unters. zu sprachl. Regeln u. ihrer Beherrschung durch Aphatiker, Kinder u. Ausländer, am Beispiel d. dt. Substantivdeklination (Vol. 82). TBL-Verlag Narr.

Rosen, E. R. (2022). Modeling human-like morphological prediction. In *Proceedings of the Society for Computation in Linguistics* (Vol. 5, pp. 133–142). doi: 10.7275/Y721-S608

Samara, A., Smith, K., Brown, H., & Wonnacott, E. (2017, May). Acquiring variation in an artificial language: Children and adults are sensitive to socially conditioned linguistic variation. *Cognitive Psychology*, *94*, 85–114. doi: 10.1016/j.cogpsych.2017.02.004

Shannon, C. E. (1948, July). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423. (Conference Name: The Bell System Technical Journal) doi: 10.1002/j.1538-7305.1948.tb01338.x

Smith, K., Perfors, A., Fehér, O., Samara, A., Swoboda, K., & Wonnacott, E. (2017, January). Language learning, language use and the evolution of linguistic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160051. doi: 10.1098/rstb.2016.0051

Smith, K., & Wonnacott, E. (2010, September). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*(3), 444–449. doi: 10.1016/j.cognition.2010.06.004

Spreng, B. (2004, January). Error patterns in the acquisition of German plural morphology: Evidence for the relevance of grammatical gender as a cue. *Toronto Working Papers in Linguistics*, *23*.

Wiese, R. (1996). *The phonology of German*. Oxford University Press on Demand.

Wiese, R. (1999, December). On default rules and other rules. *Behavioral and Brain Sciences*, *22*(6), 1043–1044. doi: 10.1017/S0140525X99532226

Williams, A., Pimentel, T., Blix, H., McCarthy, A. D., Chodroff, E., & Cotterell, R. (2020, July). Predicting Declension Class from Form and Meaning. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 6682–6695). Online: Association for Computational Linguistics.

Wu, S., Cotterell, R., & O'Donnell, T. (2019, July). Morphological Irregularity Correlates with Frequency. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 5117–5126). Florence, Italy: Association for Computational Linguistics. doi: 10.18653/v1/P19-1505

Yang, C. D. (2016). *The price of linguistic productivity : how children learn to break the rules of language*. Cambridge, Massachusetts: The MIT Press.

Zaretsky, E., & Lange, B. P. (2016). No matter how hard we try: Still no default plural marker in nonce nouns in Modern High German. In *A blend of MaLT: selected contributions from the Methods and Linguistic Theories Symposium 2015* (pp. 153–178). Bamberg: University of Bamberg Press.