The Zipfian Paradigm Cell Filling Problem

James P. Blevins[1], Petar Milin[2,3], and Michael Ramscar[3]

[1]University of Cambridge
[2]University of Sheffield
[3]Eberhard Karls Universität Tübingen

Abstract

This chapter proposes that the stable coexistence of regular and irregular patterns can be understood in terms of a trade-off between the opposing communicative pressures imposed by predictability and discriminability. On this view, irregularity is not 'defective' or 'anomalous'. Instead, irregular formations exhibit an enhanced discriminability that brings them into maximal conformance with precepts like the 'one form-one meaning principle', while allowing them to act as attractors within a larger system. Conversely, regularity is neither 'optimal' nor 'normative'. Regular patterns serve to facilitate predictability within a system. In order for regular items to perform this function, it must be possible to assign partially attested paradigms that exhaust the variation in the system. We suggest that a correlation between lexical neighbourhoods and patterns of co-filled cells bootstraps this analogical process.

1 Introduction

This chapter outlines a general view of form variation as reflecting different states of equilibrium between competing communicative pressures. Two dominant pressures are considered here. The first is a pressure to *discriminate* forms and the messages they express, which has the effect of enhancing differences between expressions. This pressure pushes forms towards the discriminative patterning expressed by the 'one form-one meaning principle'. Unchecked, this pressure can in principle lead to suppletion of the kind reported in languages such as Yélî Dnye (Henderson 1995). However, in most languages, the pressure towards maximal discrimination is

countered by a second pressure, which favours regular patterns of form and distribution that facilitate the *prediction* of unencountered forms. The trade-offs between these competing pressures helps to account for the coexistence of patterns exhibiting varying degrees of discriminability and regularity in many languages.

These pressures interact with distributional factors in characteristically different ways. Highly irregular formations, such as cases of suppletion, must maintain a sufficiently high *token frequency* in order to remain part of a language. There is likewise a *type frequency* threshold that patterns must maintain in order to qualify as regular. These requirements follow essentially a matter of definition. But regular items (i.e., members of classes with a high type frequency) also raise empirical questions about the *attestation* of variants. It has long been observed that corpora provide only partial coverage of the forms of a language, and that items with large inflectional paradigms tend to be sparsely attested. This chapter presents evidence that the shortfall is greater than previously appreciated, and that the coverage of form variation remains sparse in corpora regardless of size. Corpora obey Zipf's law at all sample sizes, containing a Large Number of Rare Events (LNRE; see Baayen 2000).

The chapter suggests that lexical neighbourhoods play a useful role in defining the classes that support extrapolation from partial samples of inflected variants. The fact that most paradigms are only partially attested inhibits the assignment of items to classes based on patterns of congruent variation across cells. Lexical neighbourhoods can help to bootstrap the process of class assignment by defining an initial clustering of items. The deductive value of this clustering derives from a strong correlation, discussed in Section 2.1 below, between similarity at the level of form within neigbourhoods and matching patterns of co-filled cells in paradigms. The co-filled cells in a class provide an analogical base for extending patterns exhibited by one member of the class to other, more sparsely attested, items. The reliability of these deductions in turn provides feedback that can guide subsequent class refinement. The challenge posed by input sparsity is met by extrapolating from classes of items that collectively exhaust the variation exhibited by the class and also contain co-filled cells that provide an analogical base for deducing unencountered forms.

1.1   The status of regularity

It is often assumed that regularity in a linguistic system is a desirable or normative state and that suppletion and other irregularities represent deviations from the uniform patterns that systems (or their speakers) strive to maintain. This conception derives in part from a pair of more general assumptions that underlie Post-Bloomfieldian approaches to linguistic analysis. The first is that recurrence entails redundancy, and the second is that structure implies decomposed representations. On this view, the goal of morphological analysis involves the distillation of patterns

into general symbolic statements, schemas or rules that describe the distribution and interpretation of isolable units of form. Descriptions that exhibit recurrent patterns are regarded as deficient on the grounds that they "fall short of scientific compactness" (Bloomfield 1933: 238) or miss "linguistically significant generalizations" (Chomsky 1965: 42ff.) that should be encapsulated in a symbolic rule system.

Both of these basic assumptions operate with notions of 'identity' and 'redundancy' that are defined primarily in terms of orthographic or phonemic transcriptions. Yet the usefulness of transcriptions for capturing the notion of 'identity' relevant for speakers is challenged by studies that probe sub-phonemic contrasts. Acoustic and psychoacoustic investigation of units ranging in size from words to segments have shown that speakers consistently produce and comprehend durational differences and other types of phonetic variation that do not determine phonemic contrasts. At the word level, Gahl (2008) and Drager (2011) found systematic differences in duration between ostensibly homophonous items in English. At the segment level, Plag et al. (2016) found "significant differences in acoustic duration between some morphemic /s/'s and /z/'s and non-morphemic /s/ and /z/".

Similar contrasts distinguish seemingly recurrent morphological units. Davis et al. (2002) reported differences in duration and fundamental frequency between words and morphologically unrelated onset words. In the domain of inflection, Baayen et al. (2003) found that speakers produced Dutch nouns with a longer mean duration when they occurred as singulars than when they occurred as the stem of the corresponding plural. Kemps et al. (2005) subsequently tested speakers' sensitivity to prosodic differences between singular nouns and 'homophonous' stems, and concluded that "acoustic differences exist between uninflected and inflected forms and that listeners are sensitive to them" (Kemps et al. 2005: 441).

In short, many apparently recurrent units are neither identical nor redundant. Hence regularity cannot be treated as an intrinsic 'design feature' of language or even as a simple reflex of a combinatoric economy metric that rewards analyses in which a recurrent element is "noted only once, with a full statement as to where it is and where it is not used" (Bloomfield 1933: 238). The pervasive regularity observed in languages cannot simply be taken for granted but stands in need of explanation.

A general explanation for regularity should not only identify the factors that are favourable to the creation of regular patterns but also clarify how the function of these patterns contributes to their persistence. At the same time, these considerations should be compatible with the ubiquity of irregularity. Exceptionless regularity is comparatively rare outside artificial languages like Esperanto and even descriptions of standard languages tend to underestimate the amount of variation.

As with regularity, a general account of irregularity should identify the sources of irregular formations and also clarify their functions. Traditions that treat regularity as normative tend to regard irregularity as functionless 'historical residue' or

'noise'. Individual approaches within these traditions develop formal strategies for accommodating deviations from regular patterns, ranging from lexical lookup routines, through 'readjustment rules' to 'default overrides'. However these strategies act essentially as correctives to a conception of language structure as normatively regular and thus offer no insight into the persistence and ubiquity of irregularity.

1.2 Discriminability and sparsity

From a discriminative learning perspective of the kind developed Ramscar & Yarlett (2007); Ramscar et al. (2010) and Ramscar (2013), the situation is reversed. Suppletive forms and other types of irregular formations tend to be maximally discriminated, establishing highly transparent form/meaning contrasts. To the extent that irregular patterns enhance the discriminability of forms, they contribute to the communicative efficiency of a language. In the discriminative model of Ramscar et al. (2013), the main difference between overtly suppletive forms such as *mouse/mice* and more regular forms such as *rat/rats* is that the former serve to accelerate the rate at which a speakers' representation of a specific form/meaning contrast becomes discriminated from the form classes that express similar contrasts. Thus learning serves to increase the general level of suppletion in form-meaning mappings.

It is thus *regularity* that stands in need of explanation in a discriminative approach. Models of grammaticalization suggest a source for regular formations in the morphologization of syntactic configurations. It remains then to account for the persistence and function of regular patterns within a morphological system. One explanation can be found in the structure of the linguistic input. Previous debates regarding the poverty of the stimulus have mainly been concerned with phenomena in the syntactic domain, where there appear to be no cases that withstand serious scrutiny (see, e.g., Pullum & Scholtz 2002; Clark & Lappin 2011). However, the problem of input sparsity arises in an acute form in the morphological domain, given the distributional biases of the forms of a language. These biases reflect Zipf's law (Zipf 1935, 1949), according to which the frequency of a word in a corpus is inversely proportional to its rank in the corpus. As language samples increase in size, they reinforce the rank-size distributions established in smaller samples.[1]

As Kurumada et al. (2013: 440) note, "while Zipfian distributions are ubiquitous across natural language, their consequences for learning are only beginning to be explored". A consequence of immediate morphological relevance is that speakers must learn a language from a partial and biased sample. Although it is generally accepted that exposure to specialized vocabulary and archaic formations may vary

---

[1]This chapter is concerned solely with the effects of the patterns described by the laws attributed to Zipf and Herdan (Herdan 1960). For discussion of the source of these patterns, see Ramscar (2017).

across individuals, it is often implicitly assumed that speakers encounter the majority of regular forms in their language. A related assumption underlies the different ways that inflectional and derivational patterns have been investigated in psycholinguistic studies. Since derivational processes are known to exhibit a high degree of item-specific variation, derivational families (de Jong et al. 2000; Mulder et al. 2014) are defined in terms of type counts. In contrast, inflectional processes are assumed to be highly productive, defining uniform paradigms within a given class. Lemma size is not expected to vary, except where forms are unavailable due to paradigm 'gaps' or 'defectiveness'. This allowed studies of inflection to abstract away from variation in type counts and focus on token counts (Baayen et al. 1997; Hay 2001).

Yet studies of corpora, which provide the best available model of language input, indicate that lemma family size also varies considerably. The linguistic significance of this variation is suggested by the fact that lemma family size is a useful predictor of derivational family size, as shown in Milin et al. (2013). But what appears to be clear in any case is that many potentially available inflected forms are unattested in corpora. As corpora increase in size, they do not converge on uniformly populated paradigms. Instead, they reinforce previously attested forms and classes while introducing progressively fewer new items. This distribution suggests that inflected variants of open-class items obey Zipf's law at all observed sample sizes.

The observation that speakers never encounter all of the inflected forms of their language entails that they must be able to solve what Ackerman et al. (2009) term the 'Paradigm Cell Filling Problem' on the basis of the forms that they do encounter. Regularity contributes to this solution by facilitating prediction from a partial and biased sample. From a learning-based perspective, regularity and irregularity are best understood in terms of the complementary functions they serve within a system. Regularity is not normative; it is merely the prerequisite for prediction from Zipf-distributed input. An increase in the discriminability of forms and contrasts aids communicative efficiency. A key point about irregular forms from this perspective is that they are *frequent* – they are in the head of the Zipfian distribution. Accordingly, they can be expected to be acquired and fully discriminated early (cf. *mouses / mice*), before the prefrontal cortex develops fully (cf. Ramscar & Gitcho 2007) so that learning is not influenced by the top-down factors that often inhibit the acquisition of irregulars in adult learners. From this perspective, irregulars would exemplify the 'end point' of language learning (as well-discriminated, suppletive forms), but because language is Zipf-discriminated, there is no 'end point' of language learning, and so the distribution needs to be predictable for the tail.

Thus the coexistence of regular patterns and irregular forms is not due solely to the inertia of functionless historical residue but reflects the interaction of competing discriminative and predictive pressures. Once established, irregular formations function as highly-discriminated exponents of properties and as attractors that en-

5

hance the salience of regular contrasts. Yet increases in the discriminability of irregulars are offset by a reduction in predictive value. The structure of the input imposes limits on how irregular a language can become without sacrificing learnability.

1.3   It takes a neighbourhood

In order for a collection of partial samples to guide the prediction of unattested forms, the forms that speakers do know must be organized in such a way that they collectively exhaust the inflectional variation in a language. This gives rise to what, adapting Ackerman et al. (2009), might be called the 'Paradigm Cell Alignment Problem' (PCAP), since speakers must recognize which sets of inflected forms can be 'pooled' to cover the variation that characterizes each class of paradigms. As clearly recognized by Hockett (1967), this problem is far more pressing for speakers than the pedagogical problem of selecting 'principal parts' or 'exemplary paradigms'.

> in his analogizing … [t]he native user of the language … operates in terms of all sorts of internally stored paradigms, many of them doubtless only partial; and he may first encounter a new basic verb in any of its inflected forms. (Hockett 1967: 221)

We suggest that the organization required to solve the PCAP is provided by lexical *neighbourhoods* (Baayen et al. 2006; Gahl et al. 2011). Neighbourhoods bootstrap the creative engine of the morphological system by seeding a class structure that permits the analogical deduction of the full system from collections of partial paradigms with systematic patterns of co-filled cells. The role assigned here to neighbourhoods thus builds on the results reported in Milin et al. (2011). In this study, analogical extrapolation from a small set of nearest neighbors allowed a system to model the choice of masculine instrumental singular allomorph by Serbian speakers presented with nonce words. More generally, regular paradigms enable speakers to generate previously unencountered forms, not by appealing to an explicit rule, or to any kind of explicit grammatical knowledge, but because they are implicit in the distribution of forms and meanings in the language as a system.

The following section outlines empirical evidence that Zipf's law determines input sparsity of inflectional variants all sample sizes and suggests how the structure of lexical neighbourhoods facilitates prediction from sparse input. A central hypothesis explored in this study is that the structure of morphological systems does not reflect the types of formal constraints on representations or rule systems proposed within theoretical models, but derives instead from three primary factors. The first is the Zipfian structure of the input that speakers are exposed to. The second is the discriminative learning strategies that speakers employ when exposed to that input. The third is the structure of lexical neighbourhoods. From this learning-based

perspective, the structure of morphological systems is not anchored in any kind of formal architecture or 'innate language faculty' but emerges mainly from the distributional biases of the forms in the system and the general-purpose discriminative learning strategies and principles of analogical deductions employed by speakers.

2   The Zipfian Paradigm Cell Filling Problem

It might seem reasonable to expect that all or most of the inflectional variants of the regular items in a language would eventually show up in the input encountered by speakers, given a large enough sample. From this perspective, one might expect an initial spike of forms with high token frequency that would gradually give way to a more uniform distribution as sample size increases. Yet this is not at all what we find if we examine the distribution of forms in corpora. Figure 1 displays rank orders in random samples of the SdeWaC corpus (Faß & Eckart 2013) of German.
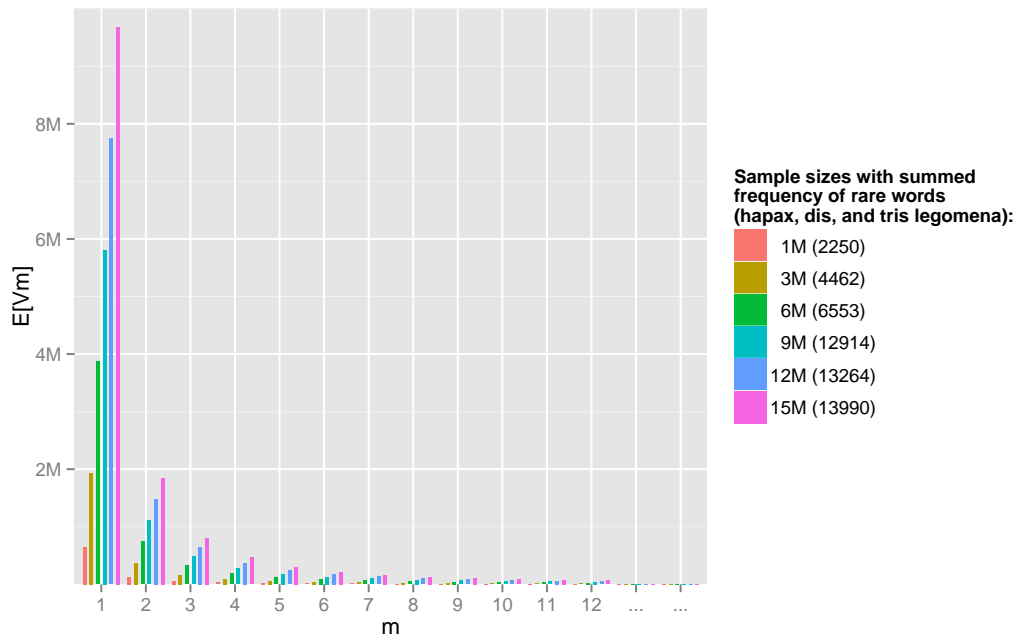

Figure 1: Zipf plot for randomly sampled words

The samples in Figure 1 start with 1 million distinct word forms and increase stepwise to to 15 million forms, at which point the 850-million word corpus is es-

sentially exhausted. At each size increment, the Zipfian structure becomes more, not less, pronounced, as the head of the distribution grows faster than its tail.

Figure 2 exhibits the result of applying a similar sampling methodology to inflected noun variants in the SdeWaC corpus. The growth in the distributional bias of nouns shows same pattern as the randomly-sampled words in Figure 1. As sample size increases, the average number of attested inflected noun variants decreases.
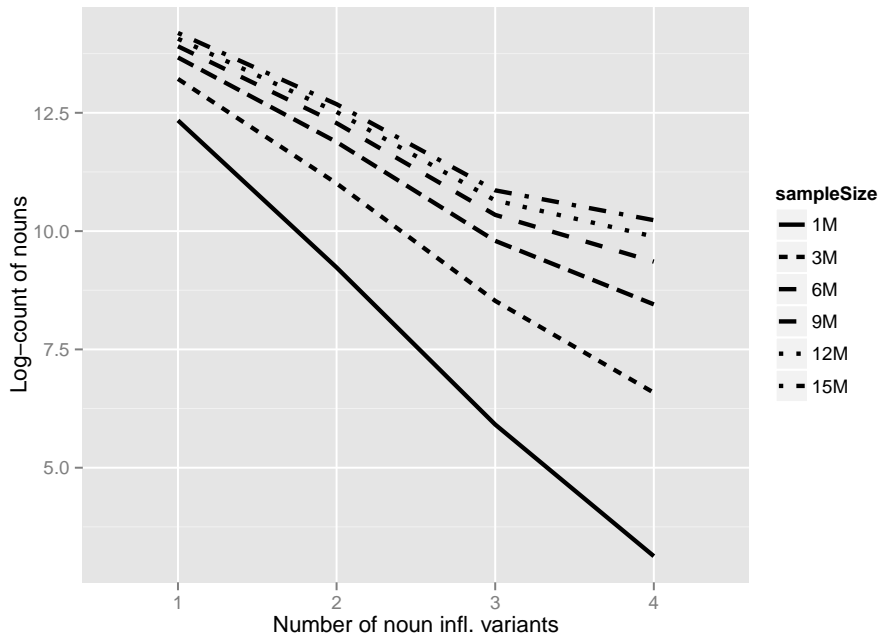

Figure 2: The paradigm non-filling pattern

As sample size increases, there is also a marked attenuation in the steepness of the slope, though it never becomes completely flat. This trend is extracted and presented in Figure 3, which plots the number of attested forms on the X-axis and slopes of six trends from Figure 2 on the Y-axis. From this relationship we can infer that even if the corpus size were increased to infinity, it would never contain all possible inflected forms of every German noun. To test this claim explicitly we applied a curve-fitting technique to estimate the slope for an unlimited corpus size ($\lim(X)$). An inverse-exponential three-parameter function obtained almost perfect fit ($R^2 = 0.996$; p-value < 0.0001). To bring even greater conservativeness, since we were extrapolating on an extremely small number of points, we conducted a numeric grid-search to find a new curve that will always behave as an upper bound

for the observed data points. This new curve is presented in Figure 3 with dashed line. With this additional restriction, the estimate for the slope when the corpus is unlimited in size is $Y = -1.326$. What this means is that no increase in input size can provide a solution to the PCFP. Even with an endless influx of new words, there would always be more nouns with a smaller number of attested inflected variants.
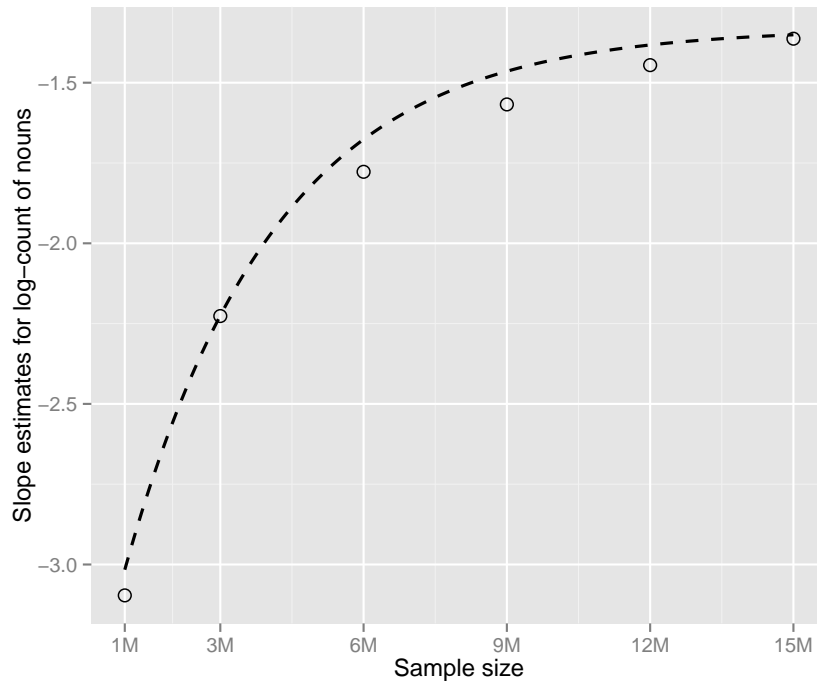

Figure 3: You can't get there from here: Asymptoting slopes

In summary, not only do the forms drawn from this corpus obey Zipf's law at all sample sizes but the rank-order biases become stronger as sample size increases. To the extent that corpora provide a reasonable approximation of input, these patterns suggest speakers must be able to acquire their language from exposure that will always be partial, and will in some respects become sparser with increased exposure.

2.1 Formal similarity and paradigmatic analogy

The Zipfian distribution exhibited in Figures 1–3 is also characteristic of the forms that fill cells of inflectional paradigms (see, e.g., Kostić et al. 2008). The exact distribution may of course be to some degree item-specific: there is no reason to assume

that the probability-based rank-order must be the same for the paradigm cells of all words. A number of different factors may influence which cells of a given paradigm are filled. Semantic considerations will play a role in some cases. For example, the physical properties of the objects associated with nouns are known to influence the probability or even possibility of particular locative or instrumental case forms.

On a Post-Bloomfieldian conception, a language is viewed as an inventory of forms composed of recurrent parts and paradigms are treated as 'epiphenomena'. Hence there are no obvious expectations about patterns of 'paradigm cell filling' apart from any effects attributable to frequency or other independent factors. In contrast, a conception of language as a complex adaptive system leads one naturally to search for functions or other correlates distributional patterns. Given that form and distribution are the two observable dimensions of variation in a morphological system, it would be surprising if there were not at least some systematic correspondences between these properties. In the present case, we consider how the shape and distribution of attested forms might aid the process of deducing unencountered forms. The specific hypothesis we explore below is that similarities in shape that define lexical neighbourhoods correlate with distributional similarities that permit attested forms to provide an analogical base for deducing unencountered forms.

This hypothesis is tested directly by evaluating whether the form proximity of wordform pairs, measured by normalized Levenshtein distance, predicts the number of co-filled paradigm cells, once effects due to the frequencies of those words are partialled out. If form similarity is a significant predictor co-filled paradigm cells, it would suggest that external considerations (such as meaning or frequency) are not the sole factors determining which cells are filled in noun paradigms in German.

2.1.1 Methodology

To test this correlation, we turned again to the SdeWaC corpus to sample German noun pairs across frequency bands. Our four-stage method represents a variant of a stratified sampling procedure in which we secured a random fraction of nouns from all frequency strata (the exact procedure was adopted from Ellis & Hooper (2001)). We first selected all nouns with a frequency of 5 or higher. We then calculated the total number of tokens (or the total summed frequency), and arbitrarily set the sample size to 50,000 nouns. The total number of tokens divided by the sample size determined the step size (i.e., the summed frequency per band). In the next stage, nouns were ordered by frequency, and ties were randomized (shuffled). In the final stage we started sampling, from low to high frequency nouns: once the cumulative sum of frequencies was equal or higher than the step size we terminated sampling for a given band and moved to the next, now using the difference between the current cumulative sum and the step size. This process was repeated until we

exhausted the full list of nouns, that is, when our sample size reached 50,000 nouns.

Once the nouns were sampled, we formed all possible pairs and calculated their normalized Levenshtein distance (Levenshtein 1966) and determined the number of co-filled paradigm cells (ranging from 0 to 8, reflecting the space defined by the 4 cases and 2 numbers in the German declension system). Since we were interested in the predictability of the closest form neighbours, we retained only those pairs whose normalized Levenshtein distance was less than or equal to 0.5 (i.e., values ranging from 0, indicating no difference, to 0.5, inidicating a difference in roughly half of the letters). Our final sample consisted of approximately 27.2 million noun pairs.

To test our prediction, we analyzed our dataset using the generalized additive mixed model (GAMM), in the R statistical environment (R Core Team 2014), with the MGCV package (Wood 2006). A GAMM was fitted to the number of co-filled cells, testing the nonlinear numeric interaction of two word frequencies and, additionally, the effect of the normalized Levenshtein distance, allowing only for mild (minimal) non-linearity. The tensor product of two frequencies appeared highly predictive (edf $=$ 23.834; F $=$ 94468; p $<$ 0.0001). Crucially, however, over and above this effect we observed a strong effect of Levenshtein distance (edf $=$ 1.986; F $=$ 3297; p $<$ 0.0001). This supports the initial hypothesis that form neighborhoods can serve as the basis for the analogical prediction of inflectional variants.

A GAMM was then fitted to the number of co-filled cells, testing the nonlinear numeric interaction of two word frequencies, cosine similarity between contextual vectors of same two words[2], and, finally, the effect of the normalized Levenshtein distance. The tensor product of two frequencies appeared highly predictive (edf $=$ 23.749; F $=$ 79595; p $<$ 0.0001), and so did cosine similarity (edf $=$ 4.000; F $=$ 138683; p $<$ 0.0001). Crucially, however, over and above these two effects we observed strong main effect of Levenshtein distance (edf $=$ 1.029; F $=$ 5805; p $<$ 0.0001). For this effect we where we allowed only for minimal nonlinearity. The partial interaction of cosine similarity with Levenshtein distance also appeared as statistically significant, but to a lesser degree (edf $=$ 15.039; F $=$ 1458; p $<$ 0.0001). These results further support the hypothesis that, in German nominal paradigms, form neighborhoods provide a basis for the analogical deduction of variants.

2.1.2 Testing the predictions

Figures 4–7 now exhibit the effects described above. Figure 4 shows that frequency effects display the expected increase in the number of co-filled cells as a product of the increase of the frequencies of the respective nouns. Figure 5 then shows the nonlinear effect of cosine similarity, revealing a general trend in which semantic

---

[2]For detailed discussion of the theoretical and technical aspects of vector-based semantic similarity, see Lund & Burgess (1996); Shaoul & Westbury (2010)
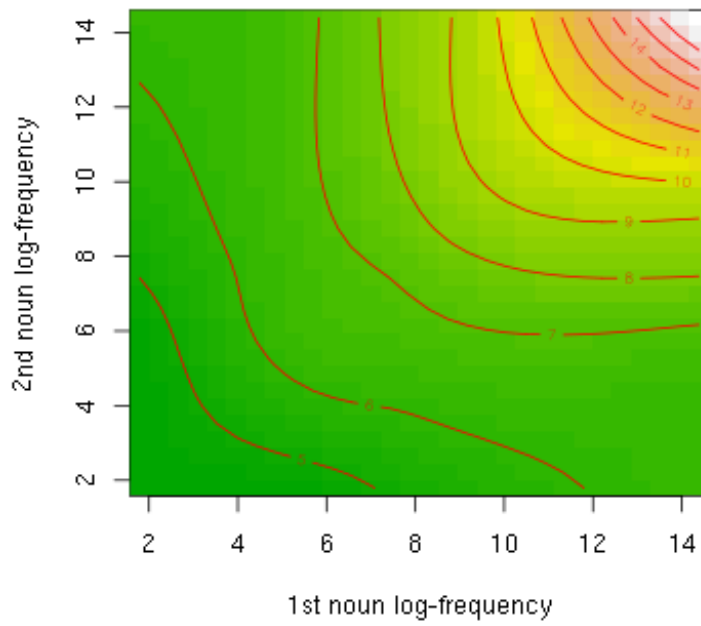
Figure 4: Nonlinear interaction (tensor product) effect of frequencies of paired nouns for the number of co-filled paradigm cells

similarity lowers the number of co-filled cells. This partial effect takes only negative values, becoming most negative for words that are most similar. The same general pattern applies to form similarity: Figure 6 shows that the number of co-filled cells steadily decreases as form similarity decreases. In this case, we see that the partial
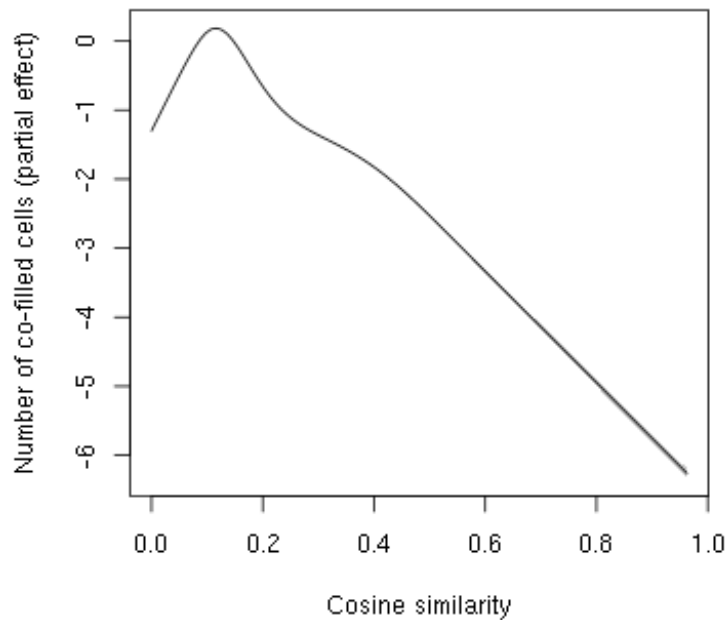

Figure 5: Partial effect of the cosine similarity between contextual vectors of two nouns for the number of co-filled paradigm cells

effect takes only positive values, but is a degree weaker than the effect of cosine similarity displayed in Figure 5.

Finally, although the partial interaction between cosine similarity and Levenshtein distance may have appeared to be statistically significant, Figure 7 reveals that it is, in fact, not substantial. Figure 7 exhibits the same cosine similarity effect, where only for the most similar nouns Levenshtein distance forms a U-shaped plane. However, we can also see that there are no nouns which are very similar in contextual appearance and closest form neighbors at the same time (that is, the upper left quadrant does not contain any values).
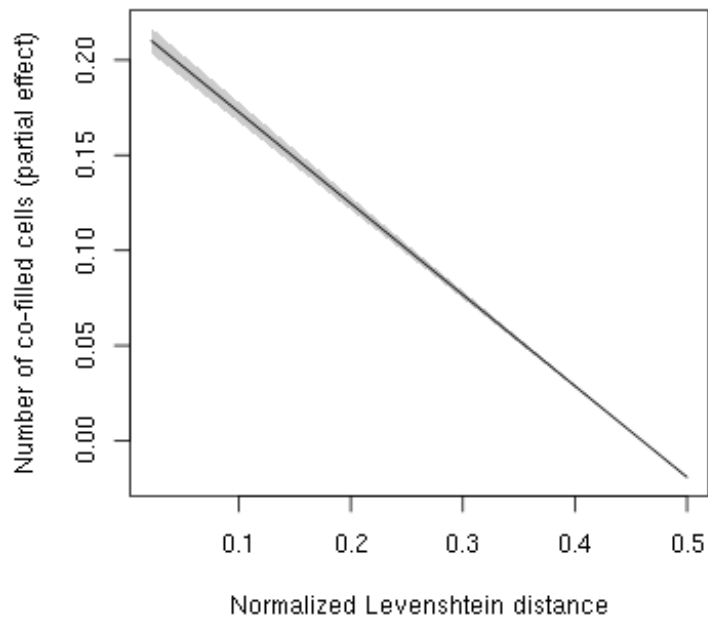
Figure 6: Partial effect of the normalized Levenshtein distance between two nouns for the number of co-filled paradigm cells
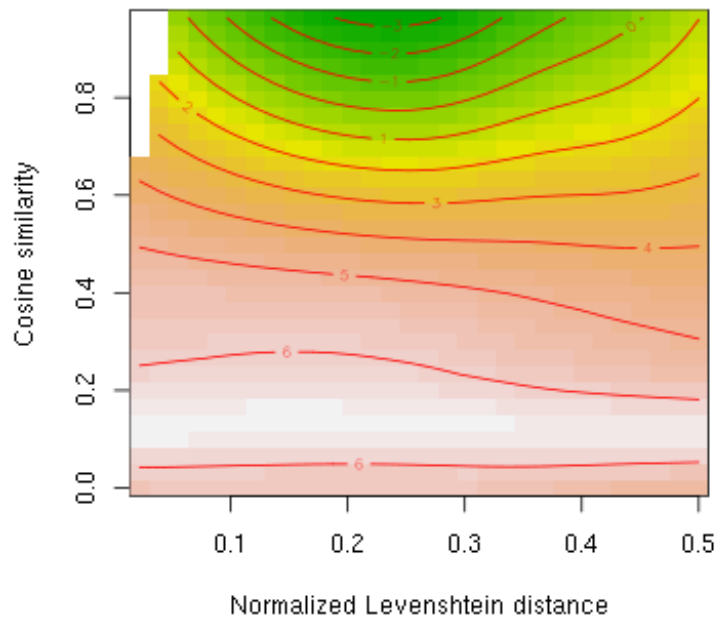
Figure 7: Nonlinear interaction (tensor product) effect of cosine similarity and Levenshtein distance between two nouns for the number of co-filled paradigm cells

2.1.3 Summary

Figures 4–7 show that the number of co-filled cells steadily decreases as the form similarity decreases. A learning framework provides a simple and elegant explanation for the observed form similarity effect (operationalized by normalized Levenshtein distance). Words attested in larger corpus should manifest analogical predictions when they are similar in form. And exactly that is confirmed by the number of co-filled paradigm cells as co-determined with the form overlap of pairs of words.

These effects reinforce the previous results of Milin et al. (2011), which showed that the closest form neighbors determine the production of unseen inflected variant of nonce words (in this case an allomorph of masculine instrumental singular). Together, these studies suggest that items similar in form indeed help paradigm cell filling, a conclusion that is also consistent with studies such as Pertsova (2004), which report neighbourhood effects on the inflection of nonce words. The present study demonstrates that the highest cell overlap characterizes the nearest neighbors.

3   A learning-based perspective on the Zipfian PCFP

We conclude with a concise summary and a brief discussion of some implications of the results reported above. The point of departure for this study was the proposal that the coexistence of regular and irregular patterns could be understood in terms of a trade-off between the opposing communicative pressures imposed by predictability and discriminability. Somewhat counterintuitively, the pressures associated with irregulars are reasonably transparent, once irregulars are recognized as functional rather than 'defective'. The enhanced discriminability of irregular formations brings them into maximal conformance with precepts like the 'one form-one meaning principle', while their effects as attractors within a larger system are established in studies such as Ramscar et al. (2013). In contrast, recognizing the role that regulars play in enhancing predictability creates a conundrum. Although, by definition, regular items must have high type frequency, the Zipfian distribution of regularly inflected variants entails that many variants have low token frequency or are unattested altogether. In effect, the forms that establish the regularity of regulars must be deduced from partial samples. Section 2.1 suggests how the relation between neighbourhoods and patterns of co-filled cells facilitate this deduction.

The elements of the resulting communicative dynamic are summarized in (1).

(1)   THE PREDICTION-DISCRIMINATION DYNAMIC

    a. Morphological systems exhibit regularities because, given the Zipfian structure of the input, speakers never encounter all the forms of a language and must be able to predict new forms from partial samples.

    b. Irregular formations are persistent because they serve two communicative functions. As individual expressions, they are well discriminated. As exceptional members of larger sets of alternating elements, they emphasize contrasts that are less saliently marked in regular patterns.

    c. Lexical neighbourhoods compensate for input sparsity. Although the forms of individual items are partially attested, the inflectional patterns that they follow are robustly attested within their form neighbourhoods.

This initial study also suggests a number of directions for future research. The relation between neighbourhoods and co-filled cells identifies a viable base for analogical extrapolation. Studies such as Ackerman & Malouf (2013) likewise show that paradigms exhibit a degree of mutual informativeness that supports the deduction of unencountered forms. A discriminative learning model is well adapted to exploiting these sources of information in language learning and processing. However, questions remain concerning the cognitive implementation of analogical deduction and the ways that this process interacts with other factors that enter into learning and processing. The amount of cross-linguistic variability in the relation between neighourhood similarity and patterns of co-filled cells also remains an open question. The initial choice of German helps to establish that input sparsity arises in languages with comparatively few inflectional variants. In languages with larger and more variable paradigms, Zipfian biases are expected to create an even greater form shortfall and thereby enhance the deductive value of an analogical base.

A clearer understanding of the way that Zipfian biases constrain the space of solutions for the PCFP is also of value to Post-Bloomfieldian models. To solve the PCFP, combinatoric approaches require a predictive model that overcomes the 'Segmentation Problem' (Spencer 2012) and other challenges to the assumption of lossless decomposition (see, e.g., Blevins 2016). Approaches that aim for psychological relevance also require a notion of 'identity' that takes account of the sub-phonemic variation described in Section 1.1, along with some evidence that the model is learnable from the Zipf-distributed input that speakers can be assumed to encounter.

References

Ackerman, F., Blevins, J. P. & Malouf, R. (2009). Parts and wholes: Implicative patterns in inflectional paradigms. In Blevins, J. P. & Blevins, J. (eds.), *Analogy in Grammar: Form and Acquisition*, Oxford University Press, chapter 3, 54–81.

Ackerman, F. & Malouf, R. (2013). Morphological organization: The Low Conditional Entropy Conjecture. *Language* **89**, 429–464.

Baayen, R. H. (2000). *Word Frequency Distributions*. Dordrecht: Kluwer Academic Publishers.

Baayen, R. H., Feldman, L. B. & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language* **53**, 496–512.

Baayen, R. H., Lieber, R. & Schreuder, R. (1997). The morphological complexity of simple nouns. *Linguistics* **35**, 861–877.

Baayen, R. H., McQueen, J. M., Dijkstra, T. & Schreuder, R. (2003). Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. In Baayen, R. H. & Schreuder, R. (eds.), *Morphological Structure in Language Processing*, Berlin: Mouton de Gruyter, 355–370.

Blevins, J. P. (2016). *Word and Paradigm Morphology*. Oxford: Oxford University Press.

Bloomfield, L. (1933). *Language*. Chicago: University of Chicago Press.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Clark, A. & Lappin, S. (2011). *Linguistic Nativism and the Poverty of the Stimulus*. Wiley-Blackwell.

Davis, M., Marslen-Wilson, W. D. & Gaskell, M. (2002). Leading up the lexical garden-path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception & Performance* **28**, 218–244.

de Jong, N. H., Schreuder, R. & Baayen, R. H. (2000). The morphological family size effect and morphology. *Language and Cognitive Processes* **15**, 329–365.

Drager, K. K. (2011). Sociophonetic variation and the lemma. *Journal of Phonetics* **39**, 694–707.

Ellis, N. C. & Hooper, A. M. (2001). Why learning to read is easier in Welsh than in English: Orthographic transparency effects evinced with frequency-matched tests. *Applied Psycholinguistics* **22(4)**, 571–599.

Faß, G. & Eckart, K. (2013). SdeWaC – a corpus of parsable sentences from the web. In Gurevych, I., Biemann, C. & Zesch, T. (eds.), *Language Processing and Knowledge in the Web*, Heidelberg: Springer, 61–68.

Gahl, S. (2008). "Thyme" and "Time" are not homophones. The effect of lemma frequency on word durations in spontaneous speech. *Language* **84(3)**, 474–496.

Gahl, S., Yao, Y. & Johnson, K. (2011). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language* **66(4)**, 789–806.

Hay, J. (2001). Lexical frequency in morphology: Is everything relative? *Linguistics* **39**, 1041–1070.

Henderson, J. E. (1995). *Phonology and Grammar of Yele, Papua New Guinea*. Pacific Linguistics B-112, Camberra: Pacific Linguistics.

Herdan, G. (1960). *Type-Token Mathematics*. The Hague: Mouton.

Hockett, C. F. (1967). The Yawelmani basic verb. *Language* **43**, 208–222.

Kemps, J. J. K., Rachèl, Ernestus, M., Schreuder, R. & Baayen, R. H. (2005). Prosodic cues for morphological complexity: The case of Dutch plural nouns. *Memory & Cognition* **33(3)**, 430–446.

Kostić, A., Ilić, S. & Milin, P. (2008). Probability estimate and the optimal text size. *Psihologija* **41(1)**, 35–51.

Kurumada, C., Meylan, S. C. & Frank, M. C. (2013). Zipfian frequency distributions facilitate word segmentation in context. *Cognition* **127**, 439–453.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady* **10(8)**, 707–710.

Lund, K. & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers* **28(2)**, 203–208.

Milin, P., Keuleers, E. & Blevins, J. P. (2013). Lemma size effects on morphological processing. 9th Mediterranean Morphology Meeting, Dubrovnik.

Milin, P., Keuleers, E. & Filipović Đurđević, D. (2011). Allomorphic responses in Serbian pseudo-nouns as a result of analogical learning. *Acta Linguistica Hungarica* **58**, 65–84.

Mulder, K., Dijkstra, T., Schreuder, R. & Baayen, R. H. (2014). Effects of primary and secondary morphological family size in monolingual and bilingual word processing. *Journal of Memory and Language* **72**, 59–84.

Pertsova, K. (2004). *Distribution of Genitive Plural Allomorphs in the Russian Lexicon and in the Internal Grammar of Native Speakers*. Master's thesis, UCLA.

Plag, I., Homann, J. & Kunter, G. (2016). Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics* **52(4)**, (in press).

Pullum, G. K. & Scholtz, B. (2002). Empirical assessment of stimulus poverty arguments. *The Linguistic Review* **19**, 8–50.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramscar, M. (2013). Suffixing, prefixing, and the functional order of regularities in meaningful strings. *Psihologija* **46(4)**, 377–396.

Ramscar, M. (2017). The discriminative nature of human communication. ms. Eberhard-Karls-Universität, Tübingen.

Ramscar, M., Dye, M. & McCauley, S. M. (2013). Error and expectation in language learning: The curious absence of *mouses* in adult speech. *Language* **89(4)**, 760–793.

Ramscar, M. & Gitcho, N. (2007). Developmental change and the nature of learning in childhood. *Trends in Cognitive Sciences* **11(7)**, 274–279.

Ramscar, M. & Yarlett, D. (2007). Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. *Cognitive Science* **31**, 927–960.

Ramscar, M., Yarlett, D., Dye, M., Denny, K. & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science* **34**, 909–957.

Shaoul, C. & Westbury, C. (2010). Exploring lexical co-occurrence space using HiDEx. *Behavior Research Methods* **42(2)**, 393–413.

Spencer, A. J. (2012). Identifying stems. *Word Structure* **5**, 88–108.

Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. CRC Press.

Zipf, G. K. (1935). *The Psychobiology of Language: An Introduction to Dynamic Philology*. Cambridge, MA: MIT Press.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley.