Much ado about nothing: A social network model of Russian paradigmatic gaps

Robert Daland **Andrea D. Sims** **Janet Pierrehumbert**

Department of Linguistics
Northwestern University
2016 Sheridan Road
Evanston, IL 60208 USA
r-daland, andrea-sims, jbp@northwestern.edu

Abstract

A number of Russian verbs lack 1sg non-past forms. These paradigmatic gaps are puzzling because they seemingly contradict the highly productive nature of inflectional systems. We model the persistence and spread of Russian gaps via a multi-agent model with Bayesian learning. We ran three simulations: no grammar learning, learning with arbitrary analogical pressure, and morphophonologically conditioned learning. We compare the results to the attested historical development of the gaps. Contradicting previous accounts, we propose that the persistence of gaps can be explained in the absence of synchronic competition between forms.

1 Introduction

Paradigmatic gaps present an interesting challenge for theories of inflectional structure and language learning. Wug tests, analogical change and children's overextensions of regular patterns demonstrate that inflectional morphology is highly productive. Yet lemmas sometimes have "missing" inflected forms. For example, in Russian the majority of verbs have first person singular (1sg) non-past forms (e.g., *posadit'* 'to plant', *posažu* 'I will plant'), but no 1sg form for a number of similar verbs (e.g., *pobedit'* 'to win', **pobežu* 'I will win'). The challenge lies in explaining this apparent contradiction. Given the highly productive nature of inflection, why do paradigmatic gaps arise? Why do they persist?

One approach explains paradigmatic gaps as a problem in generating an acceptable form. Under this hypothesis, gaps result from irreconcilable conflict between two or more inflectional patterns. For example, Albright (2003) presents an analysis of Spanish verbal gaps based on the Minimal Generalization Learner (Albright and Hayes 2002). In his account, competition between mid-vowel diphthongization (e.g., *s[e]ntir* 'to feel', *s[je]nto* 'I feel') and non-diphthongization (e.g., *p[e]dir* 'to ask', *p[i]do* 'I ask') leads to paradigmatic gaps in lexemes for which the applicability of diphthongization has low reliability (e.g., *abolir* 'to abolish, **ab[we]lo*, **ab[o]lo* 'I abolish').

However, this approach both overpredicts and underpredicts the existence of gaps cross-linguistically. First, it predicts that gaps should occur whenever the analogical forces determining word forms are contradictory and evenly weighted. However, variation between two inflectional patterns seems to more commonly result from such a scenario. Second, the model predicts that if the form-based conflict disappears, the gaps should also disappear. However, in Russian and probably in other languages, gaps persist even after the loss of competing inflectional patterns or other synchronic form-based motivation (Sims 2006).

By contrast, our approach operates at the level of inflectional property sets (IPS), or more properly, at the level of inflectional paradigms. We propose that once gaps are established in a language for whatever reason, they persist because learners infer the relative non-use of a given

Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, pages 936–943,
Prague, Czech Republic, June 2007.

combination of stem and IPS.[1] Put differently, we hypothesize that speakers possess at least two kinds of knowledge about inflectional structure: (1) knowledge of how to generate the appropriate form for a given lemma and IPS, and (2) knowledge of the probability with which that combination of lemma and property set is expressed, regardless of the form. Our approach differs from previous accounts in that persistence of gaps is attributed to the latter kind of knowledge, and does not depend on synchronic morphological competition.

We present a case study of the Russian verbal gaps, which are notable for their persistence. They arose between the mid 19th and early 20th century (Baerman 2007), and are still strongly attested in the modern language, but have no apparent synchronic morphological cause.

We model the persistence and spread of the Russian verbal gaps with a multi-agent model with Bayesian learning. Our model has two kinds of agents, adults and children. A model cycle consists of two phases: a production-perception phase, and a learning-maturation phase. In the production-perception phase, adults produce a batch of linguistic data (verb forms), and children listen to the productions from the adults they know. In the learning-maturation phase, children build a grammar based on the input they have received, then mature into adults. The existing adults die off, and the next generation of children is born.

Our model exhibits similar behavior to what is known about the development of Russian gaps.

2 The historical and distributional facts of Russian verbal gaps

2.1 Traditional descriptions

Grammars and dictionaries of Russian frequently cite paradigmatic gaps in the 1sg non-past. Nine major dictionaries and grammars, including Švedova (1982) and Zaliznjak (1977), yielded a combined list of 96 gaps representing 68 distinct stems. These verbal gaps fall almost entirely into the second conjugation class, and they overwhelmingly affect the subgroup of dental stems. Commonly cited gaps include: **galžu* 'I make a hubbub'; **očučus'* 'I come to be (REFL)'; 1SG **oščušču* 'I feel'; **pobežu* 'I will win'; and **ubežu* 'I will convince'.[2]

There is no satisfactory synchronic reason for the existence of the gaps. The grouping of gaps among 2nd conjugation dental stems is seemingly non-arbitrary because these are exactly the forms that would be subject to a palatalizing morphophonological alternation ($t^j \rightarrow$ tʃ or $ʃ^j$, $d^j \rightarrow$ ʒ, $s^j \rightarrow$ ʃ, $z^j \rightarrow$ ʒ). Yet the Russian gaps do not meet the criteria for morphophonological competition as intended by Albright's (2003) model, because the alternations apply automatically in Contemporary Standard Russian. Analogical forces should thus heavily favor a single form, for example, *pobežu*.

Traditional explanations for the gaps, such as homophony avoidance (Švedova 1982) are also unsatisfactory since they can, at best, explain only a small percentage of the gaps.

Thus, the data suggest that gaps persist in Russian primarily because they are not uttered, and this non-use is learned by succeeding generations of Russian speakers.[3] The clustering of the gaps among 2nd conjugation dental stems most likely is partially a remnant of their original causes, and partially represents analogic extension of gaps along morphophonological lines (see 2.3 below).

2.2 Empirical evidence for and operational definition of gaps

When dealing with descriptions in semi-prescriptive sources such as dictionaries, we must always ask whether they accurately represent language use. In other words, is there empirical evidence that speakers fail to use these words?

We sought evidence of gaps from the Russian National Corpus (RNC).[4] The RNC is a balanced textual corpus with 77.6 million words consisting primarily of the contemporary Russian literary language. The content is prose, plays, memoirs and biographies, literary criticism, newspaper and magazine articles, school texts, religious and

[1] Paradigmatic gaps also probably serve a sociolinguistic purpose, for example as markers of education, but socio-linguistic issues are beyond the scope of this paper.

[2] We use here the standard Cyrillic transliteration used by linguists. It should not be considered an accurate phonological representation. Elsewhere, when phonological issues are relevant, we use IPA.

[3] See Manning (2003) and Zuraw (2003) on learning from implicit negative evidence.

[4] Documentation: http://ruscorpora.ru/corpora-structure.html Mirror site used for searching: http://corpus.leeds.ac.uk/ruscorpora.html.

philosophical materials, technical and scientific texts, judicial and governmental publications, etc.

We gathered token frequencies for the six non-past forms of 3,265 randomly selected second conjugation verb lemmas. This produced 11,729 inflected forms with non-zero frequency.[5] As described in Section 3 below, these 11,729 form frequencies became our model's seed data.

To test the claim that Russian has verbal gaps, we examined a subsample of 557 2nd conjugation lemmas meeting the following criteria: (a) total non-past frequency greater than 36 raw tokens, and (b) 3sg and 3pl constituting less than 85% of total non-past frequency.[6] These constraints were designed to select verbs for which all six person-number combinations should be robustly attested, and to minimize sampling errors by removing lemmas with low attestation.

We calculated the probability of the 1sg inflection by dividing the number of 1sg forms by the total number of non-past forms. The subset was bimodally distributed with one peak near 0%, a trough at around 2%, and the other peak at 13.3%. The first peak represents lemmas in which the 1sg form is basically not used – gaps. Accordingly, we define gaps as second conjugation verbs which meet criteria (a) and (b) above, and for which the 1sg non-past form constitutes less than 2% of total non-past frequency for that lemma (N=56).

In accordance with the grammatical descriptions, our criteria are disproportionately likely to identify dental stems as gaps. Still, only 43 of 412 dental stems (10.4%) have gaps, compared with 13 gaps among 397 examples of other stems (3.3%).

Second, not all dental stems are equally affected. There seems to be a weak prototypicality effect centered around stems ending in /d^j/, from which /t^j/ and /z^j/ each differ by one phonological feature. There may also be some weak semantic factors that we do not consider here.

/$\mathbf{d}^j$/	/$\mathbf{t}^j$/	/$\mathbf{z}^j$/	/$\mathbf{s}^j$/	/$\mathbf{st}^j$/
13.3% (19/143)	12.4% (14/118)	11.9% (5/42)	4.8% (3/62)	4.3% (2/47)

Table 1. Distribution of Russian verbal gaps among dental stems

[5] We excluded 29 high-frequency lemmas for which the corpus did not provide accurate counts.

[6] Russian has a number of verbs for which only the 3sg and 3pl are regularly used.

2.3 Some relevant historical facts

A significant difference between the morphological competition approach and our statistical learning approach is that the former attempts to provide a single account for both the rise and the perpetuation of paradigmatic gaps. By contrast, our statistical learning model does not require that the morphological system provide synchronic motivation. The following question thus arises: Were the Russian gaps originally caused by forces which are no longer in play in the language?

Baerman and Corbett (2006) find evidence that the gaps began with a single root, -*bed*- (e.g., *pobedit'* 'to win'), and subsequently spread analogically within dental stems. Baerman (2007) expands on the historical evidence, finding that a conspiracy of several factors provided the initial push towards defective 1sg forms. Most important among these, many of the verbs with 1sg gaps in modern Russian are historically associated with aberrant morphophonological alternations. He argues that when these unusual alternations were eliminated in the language, some of the words failed to be integrated into the new morphological patterns, which resulted in lexically specified gaps.

Important to the point here is that the elimination of marginal alternations removed an earlier synchronic motivation for the gaps. Yet gaps have persisted and new gaps have arisen (e.g., *pylesosit'* 'to vacuum'). This persistence is the behavior that we seek to model.

3 Formal aspects of the model

We take up two questions: How much machinery do we need for gaps to persist? How much machinery do we need for gaps to spread to phonologically similar words? We model three scenarios.

In the first scenario there is no grammar learning. Adult agents produce forms by random sampling from the forms that heard as children, and child agents hear those forms. In the subsequent generation children become adults. In this scenario there is thus no analogical pressure. Any perseverance of gaps results from word-specific learning.

The second scenario is similar to the first, except that the learning process includes analogical pressure from a random set of words. Specifically, for a target concept, the estimated distribution of its IPS is influenced by the distribution of known words. This enables the learner to express a known concept with a novel IPS. For example, imagine that a learner hears the present tense verb form *googles*, but not the past tense *googled*. By analogy with other verbs, learners can expect the past tense to occur with a certain frequency, even if they have not encountered it.

The third scenario builds upon the second. In this version, the analogical pressure is not completely random. Instead, it is weighted by morphophonological similarity – similar word forms contribute more to the analogical force on a target concept than do dissimilar forms. This addition to the model is motivated by the pervasive importance of stem shape in the Russian morphological system generally, and potentially provides an account for the phonological prototypicality effect among Russian gaps.

The three scenarios thus represent increasing machinery for the model, and we use them to explore the conditions necessary for gaps to persist and spread. We created a multi-agent network model with Bayesian learning component. In the following sections we describe the model's structure, and outline the criteria by which we evaluate its output under the various conditions.

3.1 Social structure

Our model includes two generations of agents. Adult agents output linguistic forms, which provide linguistic input for child agents. Output/input occurs in batches.[7] After each batch all adults die, all children mature into adults, and a new generation of children is born. Each run of the model included 10 generations of agents.

We model the social structure with a random network. Each adult produces 100,000 verb forms, and each child is exposed to every production from every adult to whom they are connected. Each generation consisted of 50 adult agents, and child agents are connected to adults with some probability *p*. On average, each child agent is connected to 10 adult agents, meaning that each child hears, on average, 1,000,000 tokens.

3.2 Linguistic events

Russian gaps are localized to second conjugation non-past verb forms, so productions of these forms are the focus of interest. Formally, we define a linguistic event as a concept-inflection-form (C,I,F) triple. The concept serves to connect the different forms and inflections of the same lemma.

3.3 Definition of grammar

A grammar is defined as a probability distribution over linguistic events. This gives rise to natural formulations of learning and production as statistical processes: learning is estimating a probability distribution from existing data, and production is sampling from a probability distribution. The grammar can be factored into modular components:

$$p(C, I, F) = p(C) \cdot p(I \mid C) \cdot p(F \mid C, I)$$

In this paper we focus on the probability distribution of concept-inflection pairs. In other words, we focus on the relative frequency of inflectional property sets (IPS) on a lemma-by-lemma basis, represented by the middle term above.

Accordingly, we made the simplest possible assumptions for the first and last terms. To calculate the probability of a concept, children use the sample frequency (e.g., if they hear 10 tokens of the concept 'eat', and 1,000 tokens total, then p('eat') = 10/1000 = .01). Learning of forms is perfect. That is, learners always produce the correct form for every concept-inflection pair.

3.4 Learning model

Although production in the real world is governed by semantics, we treat it here as a statistical process, much like rolling a six-sided die which may or may not be fair. When producing a Russian non-past verb, there are six possible combinations of inflectional properties (3 persons * 2 numbers). In our model, word learning involves estimating the probability distribution over the frequencies of the six forms on a lemma-by-lemma basis. A hypothetical example that introduces our variables:

jest'	**1sg**	**2sg**	**3sg**	**1pl**	**2pl**	**3pl**	**SUM**
D	15	5	45	5	5	25	100
d	0.15	0.05	0.45	0.05	0.05	0.25	1

Table 2. Hypothetical probability distribution

The first row indicates the concept and the inflections. The second row (**D**) indicates the

[7] See Niyogi (2006) for why batch learning is a reasonable approximation in this context.

hypothetical number of tokens of *jest'* 'eat' that the learner heard for each inflection (bolding indicates a six-vector). We use |D| to indicate the sum of this row (=100), which is the concept frequency. The third row (**d**) indicates the sample probability of that inflection, which is simply the second row divided by |D|.

The learner's goal is to estimate the distribution that generated this data. We assume the multinomial distribution, whose parameter is simply the vector of probabilities of each IPS. For each concept, the learner's task is to estimate the probability of each IPS, represented by **h** in the equations below. We begin with Bayes' rule:

$$p(\mathbf{h} \mid \mathbf{D}) \propto p(\mathbf{h}) \cdot \text{multinom}(\mathbf{D} \mid \mathbf{h})$$

The prior distribution constitutes the analogical pressure on the lemma. It is generated from the "expected" behavior, $\mathbf{h^0}$, which is an average of the known behavior from a random sample of other lemmas. The parameter κ determines the number of lemmas that are sampled for this purpose – it represents how many existing words affect a new word. To model the effect of morphophonological similarity (mpSim), in one variant of the model we weight this average by the similarity of the stem-final consonant.[8] For example, this has the effect that existing dental stems have more of an effect on dental stems. In this case, we define

$$\mathbf{h^0} = \Sigma_{c' \text{ in sample}}\, \mathbf{d}_{c'} \cdot \text{mpSim}(c, c')/\Sigma\, \text{mpSim}(c, c')$$

We use a featural definition of similarity, so that if the stem-final consonants differ by 0, 1, 2, or 3 or more phonological features, the resulting similarity is 1, 2/3, 1/3, or 0, respectively.

The prior distribution should assign higher probability to hypotheses that are "closer" to this expected behavior $\mathbf{h^0}$. Since the hypothesis is itself a probability distribution, the natural measure to use is the KL divergence. We used an exponentially distributed prior with parameter β:

$$p(\mathbf{h}) \propto \exp(-\beta \cdot \mathbf{h^0} \,\|\, \mathbf{h})$$

[8] In Russian, the stem-final consonant is important for morphological behavior generally. Any successful Russian learner would have to extract the generalization, completely apart from the issues posed by gaps.

As will be shown shortly, β has a natural interpretation as the relative strength of the prior with respect to the observed data.

The learner calculates their final grammar by taking the mode of the posterior distribution (MAP). It can be shown that this value is given by

$$\arg\max p(\mathbf{h} \mid \mathbf{D}) = (\beta \cdot \mathbf{h^0} + |D| \cdot \mathbf{d})/(\beta + |D|)$$

Thus, the output of this learning rule is a probability vector **h** that represents the estimated probability of each of the six possible IPS's for that concept. As can be seen from the equation above, this probability vector is an average of the expected behavior $\mathbf{h^0}$ and the observed data **d**, weighted by β and the amount of observed data |D|, respectively.

Our approach entails that from the perspective of a language learner, gaps are not qualitatively distinct from productive forms. Instead, 1sg non-past gaps represent one extreme of a range of probabilities that the first person singular will be produced. In this sense, "gaps" represent an artificial boundary which we place on a gradient structure for the purpose of evaluating our model.

The contrast between our learning model and the account of gaps presented in Albright (2003) merits emphasis at this point. Generally speaking, learning a word involves at least two tasks: learning how to generate the appropriate phonological form for a given concept and inflectional property set, and learning the probability that a concept and inflectional property set will be produced at all. Albright's model focuses on the former aspect; our model focuses on the latter. In short, our account of gaps lies in the likelihood of a concept-IPS pair being expressed, not in the likelihood of a form being expressed.

3.5 Production model

We model language production as sampling from the probability distribution that is the output of the learning rule.

3.6 Seeding the model

The input to the first generation was sampled from the verbs identified in the corpus search (see 2.2). Each input set contained 1,000,000 tokens, which was the average amount of input for agents in all succeeding generations. This made the first generation's input as similar as possible to the input of all succeeding generations.

3.7 Parameter space in the three scenarios

In our model we manipulate two parameters – the strength of the analogical force on a target concept during the learning process (β), and the number of concepts which create the analogical force (κ), taken randomly from known concepts.

As discussed above, we model three scenarios. In the first scenario, there is no grammar learning, so there is only one condition (β = 0). For the second and third scenarios, we run the model with four values for β, ranging from weak to strong analogical force (0.05, 0.25, 1.25, 6.25), and two values for κ, representing influence from a small or large set of other words (30, 300).

4 Evaluating the output of the model

We evaluate the output of our model against the following question: How well do gaps persist?

We count as gaps any forms meeting the criteria outlined in 2.2 above, tabulating the number of gaps which exist for only one generation, for two total generations, etc. We define τ as the expected number of generations (out of 10) that a given concept meets the gap criteria. Thus, τ represents a gap's "life expectancy" (see Figure 1).

We found that this distribution is exponential – there are few gaps that exist for all ten generations, and lots of gaps that exist for only one, so we calculated τ with a log linear regression. Each value reported is an average over 10 runs.

As discussed above, our goal was to discover whether the model can exhibit the same qualitative behavior as the historical development of Russian gaps. Persistence across a handful of generations (so far) and spread to a limited number of similar forms should be reflected by a non-negligible τ.

5 Results

In this section we present the results of our model under the scenarios and parameter settings above.

Remember that in the first scenario there is no grammar learning. This run of the model represents the baseline condition – completely word-specific knowledge. Sampling results in random walks on form frequencies, so once a word form disappears it never returns to the sample. Word-specific learning is thus sufficient for the perseverance of existing paradigmatic gaps and the creation of new ones. With no analogical pressure, gaps are robustly attested (τ = 6.32). However, the new gaps are not restricted to the 1sg, and under this scenario, learners are unable to generalize to a novel pairing of lexeme + IPS.

The second scenario presents a more complicated picture. As shown in Table 3, as analogical pressure (β) increases, gap life expectancy (τ) decreases. In other words, high analogical pressure quickly eliminates atypical frequency distributions, such as those exhibited by gaps. The runs with low values of β are particularly interesting because they represent an approximate balance between elimination of gaps as a general behavior, and the short-term persistence and even spread of gaps due to sampling artifacts and the influence of existing gaps. Thus, although the limit behavior is for gaps to disappear, this scenario retains the ability to explain persistence of gaps due to word-specific learning when there is weak analogical force.

At the same time, the facts of Russian differ from the behavior of the model in that the Russian gaps spread to morphophonologically similar forms, not random ones. The third version of our model weights the analogical strength of different concepts based upon morphophonological similarity to the target.

κ	β	τ (random)	τ (phono.)
--	0	6.32	
30	0.05	4.95	5.77
30	0.25	3.46	5.28
30	1.25	1.91	3.07
30	6.25	2.59	1.87
300	0.05	4.97	5.99
300	0.25	3.72	5.14
300	1.25	1.90	3.10
300	6.25	2.62	1.84

Table 3. Life expectancy of gaps, as a function of the strength of random analogical forces

Under these conditions we get two interesting results, presented in Table 3 above. First, gaps persist slightly better overall in scenario 3 than in scenario 2 for all levels of κ and β.[9] Compare the τ values for random analogical force (scenario 2) with the τ values for morphophonologically weighted analogical force (scenario 3).

Second, strength of analogical force matters. When there is weak analogical pressure, weighting for morphophonological similarity has little effect on the persistence and spread of gaps. However, when there is relatively strong analogical pressure, morphophonological similarity helps atypical frequency distributions to persist, as shown in Figure 1. This results from the fact that there is a prototypicality effect for gaps. Since dental stems are more likely to be gaps, incorporating sensitivity to stem shape causes the analogical pressure on target dental stems to be relatively stronger from words that are gaps. Correspondingly, the analogical pressure on non-dental stems is relatively stronger from words that are not gaps. The prototypical stem shape for a gap is thereby perpetuated and gaps spread to new dental stems.

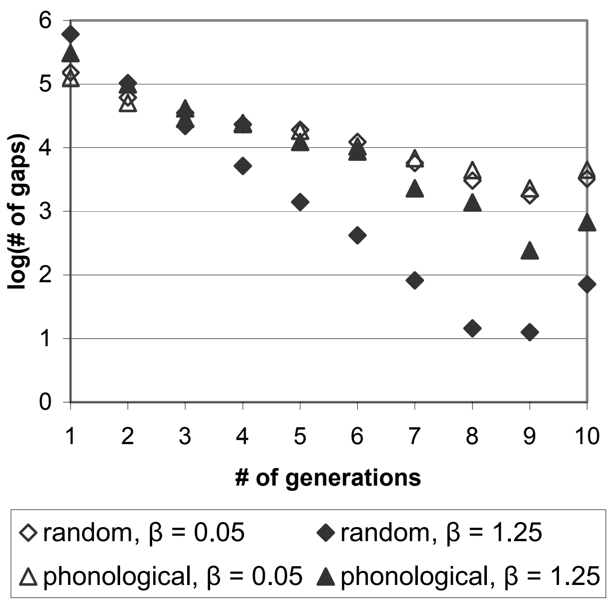

Figure 1. Gap life expectancy (β=0.05, κ=30)

[9] The apparent increase in gap half-life when β=6.25 is an artifact of the regression model. There were a few well-entrenched gaps whose high lemma frequency enables them to resist even high levels of analogical pressure over 10 generations. These data points skewed the regression, as shown by a much lower R^2 (0.5 vs. 0.85 or higher for all the other conditions).

6 Discussion

In conclusion, our model has in many respects succeeded in getting gaps to perpetuate and spread. With word-specific learning alone, well-entrenched gaps can be maintained across multiple generations. More significantly, weak analogical pressure, especially if weighted for morpho-phonological similarity, results in the perseverance and short-term growth of gaps. This is essentially the historical pattern of the Russian verbal gaps.

These results highlight several issues regarding both the nature of paradigmatic gaps and the structure of inflectional systems generally.

We claim that it is not necessary to posit an irreconcilable conflict in the generation of inflected forms in order to account for gaps. Remember that in our model, agents face no conflict in terms of which form to produce – there is only one possibility. Yet the gaps persist in part because of analogical pressure from existing gaps. Albright (2003) himself is agnostic on the issue of whether form-based competition is *necessary* for the existence and persistence of gaps, but Hudson (2000), among others, claims that gaps could not exist in the absence of it. We have presented evidence that this claim is unfounded.

But why would someone assume that grammar competition is necessary? Hudson's claim arises from a confusion of two issues. Discussing the English paradigmatic gap *amn't*, Hudson states that "a simple application of [the usage-based learning] principle would be to say that the gap exists simply because nobody says *amn't*... But this explanation is too simple... There are many inflected words that may never have been uttered, but which we can nevertheless imagine ourselves using, given the need; we generate them by generalization" (Hudson 2000:300). By his logic, there must therefore be some source of grammar conflict which prevents speakers from generalizing.

However, there is a substantial difference between having no information about a word, and having information about the non-usage of a word. We do not dispute learners' ability to generalize. We only claim that information of non-usage is sufficient to block such generalizations. When confronted with a new word, speakers will happily generalize a word form, but this is not the same task that they perform when faced with gaps.

The perseverance of gaps in the absence of form-based competition shows that a different, non-form level of representation is at issue. Generating inflectional morphology involves at least two different types of knowledge: knowledge about the appropriate word form to express a given concept and IPS on the one hand, and knowledge of how often that concept and IPS is expressed on the other. The emergence of paradigmatic gaps may be closely tied to the first type of knowledge, but the Russian gaps, at least, persist because of the second type of knowledge. We therefore propose that morphology may be defective at the morphosyntactic level.

This returns us to the question that we began this paper with – how paradigmatic gaps can persist in light of the overwhelming productivity of inflectional morphology. Our model suggests that the apparent contradiction is, at least in some cases, illusory. Productivity refers to the likelihood of a given inflectional pattern applying to a given combination of stem and IPS. Our account is based in the likelihood of the stem and inflectional property set being expressed at all, regardless of the form. In short, the Russian paradigmatic gaps represent an issue which is orthogonal to productivity. The two issues are easily confused, however. An unusual frequency distribution can make it *appear* that there is in fact a problem at the level of form, even when there may not be.

Finally, our simulations raise the question of whether the 1sg non-past gaps in Russian will persist in the language in the long term. In our model, analogical forces delay convergence to the mean, but the limit behavior is that all gaps disappear. Although there is evidence in Russian that words can develop new gaps, we do not know with any great accuracy whether the set of gaps is currently expanding, contracting, or approximately stable. Our model predicts that in the long run, the gaps will disappear under general analogical pressure. However, another possibility is that our model includes only enough factors (e.g., morphophonological similarity) to approximate the short-term influences on the Russian gaps and that we would need more factors, such as semantics, to successfully model their long-term development. This remains an open question.

References

Albright, Adam. 2003. A quantitative study of Spanish paradigm gaps. In *West Coast Conference on Formal Linguistics 22 proceedings*, eds. Gina Garding and Mimu Tsujimura. Somerville, MA: Cascadilla Press, 1-14.

Albright, Adam, and Bruce Hayes. 2002. Modeling English past tense intuitions with minimal generalization. In *Proceedings of the Sixth Meeting of the Association for Computational Linguistics Special Interest Group in Computational Phonology in Philadelphia, July 2002*, ed. Michael Maxwell. Cambridge, MA: Association for Computational Linguistics, 58-69.

Baerman, Matthew. 2007. The diachrony of defectiveness. Paper presented at 43rd Annual Meeting of the Chicago Linguistic Society in Chicago, IL, May 3-5, 2007.

Baerman, Matthew, and Greville Corbett. 2006. Three types of defective paradigms. Paper presented at The Annual Meeting of the Linguistic Society of America in Albuquerque, NM, January 5-8, 2006.

Hudson, Richard. 2000. *I amn't. *Language* 76 (2):297-323.

Manning, Christopher. 2003. Probabilistic syntax. In *Probabilistic linguistics*, eds. Rens Bod, Jennifer Hay and Stephanie Jannedy. Cambridge, MA: MIT Press, 289-341.

Niyogi, Partha. 2006. *The computational nature of language learning and evolution*. Cambridge, MA: MIT Press.

Sims, Andrea. 2006. *Minding the gaps: Inflectional defectiveness in paradigmatic morphology*. Ph.D. thesis: Linguistics Department, The Ohio State University.

Švedova, Julja. 1982. *Grammatika sovremennogo russkogo literaturnogo jayzka*. Moscow: Nauka.

Zaliznjak, A.A., ed. 1977. *Grammatičeskij slovar' russkogo jazyka: Slovoizmenenie*. Moskva: Russkij jazyk.

Zuraw, Kie. 2003. Probability in language change. In *Probabilistic linguistics*, eds. Rens Bod, Jennifer Hay and Stephanie Jannedy. Cambridge, MA: MIT Press, 139-176.