What statistics do learners track?
Rules, constraints and schemas in (artificial) grammar learning

Vsevolod Kapatsinski
University of Oregon

1. Introduction

1.1. Theoretical background
All theories of grammar specify the types of generalizations that a human language user relies
on in using language productively and thus restrict the human language learner to pay attention
only to certain types of patterns in the data to which s/he is exposed. For instance, Chomsky
and Halle (1968) and Albright and Hayes (2003), among others, assume reliance on rules. By
contrast, Bybee (2001:128) writes that

[R]ules express source-oriented generalizations. That is, they act on a specific
input to change it in well-defined ways into an output of a certain form. Many, if
not all, schemas are product-oriented rather than source-oriented. A product-
oriented schema generalizes over forms of a specific category, but does not
specify how to derive that category from some other.

The importance of product-oriented generalizations was first influentially pointed out by
Kisseberth (1970), who noted that rules conspire to produce certain types of outputs and avoid
others. This observation led to a paradigm shift in phonology from generative rules (Chomsky
and Halle 1968) towards Optimality Theory (Prince and Smolensky 1993/2004), which allows
explicit encoding of a particular kind of product-oriented generalizations (markedness
constraints) in the grammar.

By specifying the types of generalizations that can be part of a human language learner’s
grammar, theories of grammar propose the existence of a hard formal bias on learning, which
predisposes the learner to acquire specific types of generalizations and not to acquire others (or,
perhaps, to rely on only some types of generalizations that have been acquired in using the
language productively). While some theories (in particular, Optimality Theory) constrain the
learner even further by endowing him/her with an innate set of generalizations, even theories
that do not make this claim (e.g., Bybee 1985, 2001, or Albright and Hayes 2003) assume that
the grammar contains only certain types of generalizations. Optimality Theory (along with rules-plus-constraints approaches like Blevins 1997, Paradis 1989, and Roca 1997) assumes that the learner relies on a system combining both product-oriented generalizations (markedness constraints) and source-oriented generalizations (faithfulness or paradigm uniformity constraints). Network Theory (Bybee 1985, 2001), as the quote above indicates, raises the possibility of a completely product-oriented grammar. Finally, the Minimal Generalization Learner (Albright and Hayes 2003) learns only source-oriented generalizations.

1.2. Prior empirical work

The present article tests whether (adult) learners have a bias in favor of either source-oriented or product-oriented generalizations. Much of the prior experimental evidence for product-oriented generalizations is summarized in Bybee (2001:126-129). In most previous studies (Bybee and Slobin 1982, Bybee and Moder 1983, Köpcke 1988, Lobben 1991, Wang and Derwing 1994, Albright and Hayes 2003), the argument for product-oriented generalizations rests on finding that instead of respecting the input-output mappings present in the lexicon, subjects ‘overuse’ common output patterns deriving them in ways not attested in the lexicon. Unfortunately, the overuse can also be explained by experiment-internal response priming (cf. also Bickel et al. 2007, Caballero in press).

More evidence for product-oriented generalizations in natural languages is provided by cases of echolalia, in which a morpheme is not attached to a form if the form sounds like it already has the morpheme (Menn and MacWhinney 1984, Stemberger 1981, Bybee 2001:128). For instance, *It was thundering and lightning*, not *It was thundering and *lightninging*. Here speakers of English appear to be using the generalization that the progressive should end in -ing, not that one should add –ing to form the progressive. The stability of the no-change class of English verbs and its apparent resistance to overgeneralization is another possible example of this phenomenon (Menn and McWhinney 1984, Stemberger 1981, Bybee 2001:128). Phonological factors and checking of the output after the application of the –ing-adding rule (Pinker 1999:61-62) are possible alternative explanations.

Finally, evidence in favor of product-oriented generalizations is provided by Becker and Fainleib (2009) who report a miniature artificial language experiment with native Hebrew speakers. Hebrew prefers to attach the plural –ot, rather than –im to singulars whose last vowel is [o] but –im to singulars whose final vowel is [i], resulting in oCot and iCim plurals. Becker and Fainleib exposed Hebrew speakers to one of two artificial languages, which the subjects
were told were “new kinds of Hebrew”. In the “surface” pseudo-Hebrew, iC-final singulars corresponded to oCot-final plurals, while oC-final singulars corresponded to iCim-final plurals. Thus, the language users could transfer product-oriented generalizations from their native language into the artificial language. In the “deep” pseudo-Hebrew, iC-final singulars corresponded to oCim-final plurals, while oC-final singulars corresponded to iCot-final plurals. Thus, the plural forms did not obey the product-oriented generalizations that could be made on the basis of Hebrew. On the other hand, the singular-plural mappings obeyed the Hebrew rule that –ot was to be added to oC-final singulars while –im was to be added to iC-final singulars. The Hebrew learners found the surface pseudo-Hebrew easier to learn than the deep pseudo-Hebrew, suggesting that the product-oriented patterns of Hebrew transferred into the pseudo-Hebrew more easily than the source-oriented patterns, in turn suggesting that Hebrew speakers rely on product-oriented generalizations in plural formation. Becker and Fainleib hypothesize that the product-oriented generalizations are negative product-oriented generalizations that are combined with source-oriented paradigm uniformity constraints in accordance with Optimality Theory.

However, as Becker and Fainleib’s (2009) own simulations show, the source-oriented Minimal Generalization Learner (Albright and Hayes 2003) pretrained on Hebrew achieves equal accuracy on both types of pseudo-Hebrew, rather than achieving higher accuracy on deep pseudo-Hebrew. This happens because Hebrew does not feature the singular-plural mappings oC→iCot and iC→oCim found in the deep pseudo-Hebrew. Thus, unless one forces the source-oriented learner to treat the singular-plural mapping as a two-stage process, with one stage selecting the affix and relying on rules shared between Hebrew and deep pseudo-Hebrew, and the other changing the vowel (which is only needed in pseudo-Hebrew), the learner will not rely on the same source-oriented generalizations in Hebrew and deep pseudo-Hebrew. Thus, Becker and Fainleib’s results are open to the interpretation that Hebrew plural formation relies largely on source-oriented generalizations, and that product-oriented generalizations are used only when source-oriented generalizations are inapplicable.

1.3. The present experiment
The present experiment exposes native speakers of English to artificial languages that feature a process of velar palatalization before the plural suffix -i (k→tʃi) but differ in whether –i is also shown to attach to [tʃ]. Examples of tʃ→tʃi exemplify both the product-oriented generalization ‘plurals often end in -tʃi’, which favors mapping any source (including one ending in [k]) onto [tʃi], and the source-oriented generalizations ‘0 → i / C’ (which is
extracted from the same data by the Minimal Generalization Learner, developed by Albright & Hayes 2003) and ‘the stem-final consonant is retained in the plural form’ (which is predicted to be active in all languages by Optimality Theory). Thus, if typical characteristics of source-product mapping are more salient than typical characteristics of product forms, examples of /tʃ/ → /ʃi/ should disfavor palatalization, i.e., the addition of such examples to training should favor {k;t;p} → {k;t;p}i over {k;t;p} → /ʃi/. On the other hand, if product characteristics are more salient than characteristics of source-product mappings, the same examples should favor palatalization, i.e., the addition of such examples to training should favor {k;t;p} → /ʃi/. Furthermore, across subjects, source-product mappings produced by the same generalization (whether product-oriented or source-oriented) should correlate in productivity. Thus, if examples of /tʃ/ → /ʃi/ primarily exemplify X → /ʃi/ rather than 0 → /ʃi/ for a given subject should correlate with the productivity of {k;t;p} → /ʃi/ for the same subject more than with the productivity of {k;t;p} → {k;t;p}i.

Two different training paradigms were used in the present study. In source-oriented training, the artificial languages are learned under presentation conditions that can be argued to be maximally favorable for noticing relationships between source and product forms: learners are asked to repeat source-product pairs and tested on forming the product when presented with a source form. If typical characteristics of products are more noticeable than typical characteristics of source-product mappings even in this experimental paradigm, resulting in the formation and use of product-oriented generalizations, we would have strong evidence for language learners having a bias in favor of product-oriented generalizations. In product-oriented training, source and product forms sharing the same stem are no longer adjacent, with all wordforms being presented in random order. Comparison of the subjects’ behavior following two types of training can shed light on whether the extent to which learners of a language rely on product-oriented vs. source-oriented generalizations depends on the conditions under which the language is presented.

2. Methods
2.1. Languages
2.1.1. The source-oriented paradigm
A given learner was exposed to one of the languages shown in Table 1. Both languages had 30 singular-plural pairs illustrating velar palatalization. Language 1 had no singulars ending in an alveopalatal, while Language 2 had 20 singular-plural pairs featuring such a singular. Each singular-plural pair was presented twice during training. The large number of different word
types that are presented to subjects and the low token/type ratio are expected to result in
generalization across words and lack of memorization of individual wordforms. This feature of
the present training paradigm is distinct from the product-oriented paradigm, where subjects are
presented with a relatively small number of frequently occurring words that they are asked to
 memorize.

Table 1: The languages presented to learners in the source-oriented paradigm

<table>
<thead>
<tr>
<th>Language 1</th>
<th>Language 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>{t;d} \rightarrow {t;d}i</td>
<td>0</td>
</tr>
<tr>
<td>{k:g} \rightarrow {t;d}i</td>
<td></td>
</tr>
<tr>
<td>{t;d;p;b} \rightarrow {t;d;p;b}i</td>
<td>16(^1)</td>
</tr>
<tr>
<td>{t;d;p;b} \rightarrow {t;d;p;b}a</td>
<td>16(^1)</td>
</tr>
</tbody>
</table>

2.1.2. The product-oriented paradigm
In the product-oriented paradigm, subjects were exposed to individual singular and plural forms
in random order. The number of distinct words had to be reduced in order for the subjects to be
able to notice the relationship between the two forms of a given word within the same
timeframe. The languages are shown in Table 2.

Table 2: The languages presented to learners in product-oriented training

<table>
<thead>
<tr>
<th>Language 1</th>
<th>Language 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>t(\rightarrow) t(\rightarrow)i</td>
<td>0</td>
</tr>
<tr>
<td>k \rightarrow t(\rightarrow)i</td>
<td></td>
</tr>
<tr>
<td>{t;p} \rightarrow {t;p}i</td>
<td>4(^2)</td>
</tr>
<tr>
<td>{t;p} \rightarrow {t;p}a</td>
<td>4(^2)</td>
</tr>
</tbody>
</table>

Goldberg, Casenhiser and Sethuraman (2004) have shown that the learning of novel argument
structure constructions is facilitated if a few of the verbs associated with a construction occur
very often while the majority occur infrequently compared to a condition in which all verbs
occur equally often. Goldberg (2006:85-89) reports that the same result also holds for dot
pattern classification, suggesting that it is not a peculiarity of syntax (where the meaning of the

1 Half of the subjects were exposed to 24 words taking –i and 8 words taking -a while the other half were exposed to
the reversed proportions. See Kapatsinski (in press) for the significance and results of this manipulation.
2 Half of the subjects were exposed to 6 words taking –i and 2 words taking -a while the other half were exposed to
the reversed proportions.
construction might be gleaned off the meaning of the most frequent verb) and thus may also hold for morphophonology. Therefore, one word exemplifying k→tʃi, one word exemplifying the most frequent p→pV pattern in each language, one word exemplifying the most frequent t→tV pattern in each language, and one word exemplifying tʃ→tʃi were presented 42 times each, while the other words were presented 14 times each.

Recchia, Johns and Jones (2008) exposed human learners to an artificial lexicon in which words differed in frequency and the number of different sentences and pictorial scenes they appeared in. They found that frequency of presentation influenced lexical decision only if the word appeared in multiple different contexts, i.e., it had high contextual diversity. Contextual diversity was increased in the present experiment by combining each word with multiple frames: each word could be inserted in the sentences ‘{That’s a; Those are the} ___’ and ‘{I am a; We are the} ___’, and also appeared on its own and produced in a scared voice, a normal voice, or a touched voice. In addition, a voice was created for each individual creature by manipulating the speed, shifting the formant ratio, the pitch median, and the pitch range of the original speaker (me) using the ‘Change gender’ function in Praat (Boersma and Weenink 2009). The individual creature voices were used for producing the utterances fitting the schema ‘{I am a; We are the} ___’. In addition, for the frequent words, the isolated word productions were produced in four different creature voices each.

2.2. Tasks
2.2.1. The source-oriented paradigm
The experiment consisted of a training stage, an elicited production test, and a likelihood rating test. During training, participants were asked to learn “how to form plurals in the language”. A participant would be presented with a series of trials, each of which began with the presentation of a picture of a novel object on the computer screen. Three hundred milliseconds later, the name of the novel object in one of the four artificial languages was presented auditorily over headphones. Once the sound finished playing, the picture was removed and replaced with a picture of multiple (5-8) objects of the same type. The picture of multiple objects was accompanied by the auditory presentation of the plural form of the previously presented noun. Once the sound file finished playing, the participant repeated the singular-plural pair and clicked a mouse button to continue to the next singular-plural pair. The training task is shown schematically in Figure 1.
The training stage was followed by the elicited production test, which was exactly like training except instead of hearing the plural form and repeating the singular-plural pair, the learner had to generate the plural and pronounce it aloud. Half of the singulars presented during the testing were novel, i.e., they have not been presented during training. The learner was not required to repeat the singular during the test. The task is shown schematically in Figure 2.

The elicited production test was followed by the rating task. In the rating task, the subject was presented with a singular-plural pair as s/he would be during training and had to answer “How likely is this plural to be the right plural for this singular?” on a scale from 1 = ”impossible” to 5 = ”very likely”. The scale was displayed on the screen, and the learner responded by clicking a numbered rectangle with the mouse. All of the singular-plural pairs were novel and were presented in random order. Examples of the following mappings were presented for rating:
The task is presented schematically in Figure 3.

**Figure 3:** The ratings task.

<table>
<thead>
<tr>
<th>Video:</th>
<th><img src="image1.png" alt="Image" /></th>
<th><img src="image2.png" alt="Image" /></th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio:</td>
<td>[fruk]</td>
<td>[fruki]</td>
</tr>
<tr>
<td>Learner action:</td>
<td>Watch</td>
<td>Watch and listen</td>
</tr>
<tr>
<td>Duration:</td>
<td>300 ms</td>
<td>500-900 ms</td>
</tr>
</tbody>
</table>

2.2.2. The product-oriented paradigm

Like in the source-oriented training paradigm, each singular-plural pair was matched with a picture pair. However, pairings of singular nouns with objects and pairings of plural nouns with objects appeared in random order. The learner was asked to learn the names for the objects. The learner repeated the noun forms they were presented with. If the noun appeared in a sentential frame, only the noun needed to be repeated. The training task is shown schematically in Figure 4.

**Figure 4:** The product-oriented training task

<table>
<thead>
<tr>
<th>Video:</th>
<th><img src="image3.png" alt="Image" /></th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio:</td>
<td>[bootʃi]</td>
</tr>
<tr>
<td>Learner action:</td>
<td>Watch</td>
</tr>
<tr>
<td>Duration:</td>
<td>600 ms</td>
</tr>
</tbody>
</table>

After going through the training set once, the learners were tested on recalling the object names by being presented with an object or a set of identical objects and asked for the corresponding noun form. They were instructed to produce the right form of the noun (whether singular or
plural). The training-recall sequence was repeated twice and then followed by the same generalization and rating tasks used in the source-oriented paradigm.

2.3. Stimulus recording

The auditory stimuli were recorded by the author in a sound-attenuated booth onto a computer. The stimuli were sampled at 44.1 kHz and leveled to have the same mean amplitude. They were presented to the learners at a comfortable listening level of 63 dB. The learners were asked to repeat words they are hearing during training immediately after hearing them. Repetition accuracy was very high (97%). The visual stimuli were a set of made-up creature pictures retrieved from the website http://www.spore.com/sporepedia and are exemplified in Figures 1-4. The number of creatures paired with a plural wordform varied between 5 and 8. All pictures were presented on a black background.

2.4. Procedures

Learners were tested one a time. The learner was seated in a sound-attenuated booth. The audio stimuli were delivered and the learners’ speech recorded using a Sennheiser HMD281 headset. The experimenter was seated outside the booth and was able to hear the audio presented to the learner as well as the learner’s productions. The learner was unable to see the experimenter. The subject’s productions were scored by the experimenter online, as the learner was producing them. The stimuli were presented and ratings recorded using PsyScript experiment presentation software on Mac OS9.2. The order of presentation of the stimuli was randomized separately for each learner.

2.5. Participants

Participants were assigned to languages in the order they came in (Subject 1 – Language 1, Subject 8 – Language 2, etc.) In the source-oriented paradigm, 22 participants were exposed to each language. Each participant was exposed to only one language. In the product-oriented paradigm, there were also 22 participants assigned to learn each language. However, one participant assigned to Language 2 was subsequently excluded because of forming plurals using a pattern that was not presented in training (adding [tʃa]). One participant assigned to Language 1 was excluded from analyses of ratings because of computer error resulting in his ratings being lost. All of the participants reported being native English speakers with no history of speech, language, or hearing impairments. None reported being fluent in a foreign language. The participants were recruited from introductory psychology classes and received course credit for participation.
2.6. Analyses

All statistical analyses were conducted in R (http://www.cran.r-project.org). Due to severe non-normality of the data distributions, non-parametric statistics were used, i.e., all numerical variables were rank-transformed for the purposes of significance testing. The clustering solution is based on the coordinate matrix of the output of principal components analysis done on the correlation matrix between individual subjects’ production probabilities and mean ratings of examples of source-product mappings (with one point per mapping per modality per subject) with centering and scaling. The coordinate matrix contains the locations of various mappings in the multidimensional space defined by the principal components, which are orthogonal dimensions that together accounted for between-subject variance. Clustering was done using Manhattan distance, since subjects are independent non-interacting dimensions, and the Average clustering method; Ward clustering, McQuitty clustering, and Complete clustering yield the same solution.

3. Results

Figure 5 shows a hierarchical clustering solution for correlations of all mappings used or rated in production and perception following source-oriented training. The basic logic of this analysis is that if the same generalization underlies two source-product mappings, then subjects who assign a high weight to the generalization should consider both mappings acceptable, and subjects who assign a low weight to the generalization should consider both mappings unacceptable. Thus, we should find that the subjects’ ratings and production probabilities for mappings that are produced by the same generalization should show a positive correlation, and those that are produced by different generalizations should not. In this graph, the further to the right the vertical connection between two singular-plural mappings, the less similarly they were treated by the subjects, i.e., the further from 1 and the closer to -1 the correlation (r) between the mappings. In the interest of space I am omitting the very similar clustering solution for product-oriented training (the same clusters are formed at the top two branching levels).
Figure 5: The clustering of the correlation matrix between ratings and production probabilities of various mappings following source-oriented training. ‘R’ stands for ‘ratings’, while ‘P’ stands for production probabilities.

<table>
<thead>
<tr>
<th>Source</th>
<th>Product</th>
<th>Modality</th>
<th>Distance (standardized)</th>
</tr>
</thead>
<tbody>
<tr>
<td>k</td>
<td>tʃi</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>tʃi</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>tʃ</td>
<td>tʃi</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>tʃi</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>tʃi</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>tʃi</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>tʃ</td>
<td>tʃi</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>ti</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>ki</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>tʃ</td>
<td>ki</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>ti</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>ki</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>i̯</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>tʃa</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>ta</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>tʃ</td>
<td>ka</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>tʃa</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>ka</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>tʃ</td>
<td>tʃa</td>
<td>R</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5 shows that even source-oriented training results in a cluster of source-product mappings in which any source is mapped onto [tʃi], a cluster in which –i is simply added to the singular, and a cluster in which any source is mapped onto a product ending on [a]. Thus even following source-oriented training, [tʃ]→[tʃi] is unambiguously classified as an instance of C→[tʃi] rather than an instance of ‘just add –i’ in both perception and production. These results provide support for the overall primacy of product-oriented generalizations over source-oriented generalizations (Bybee 2001) and suggest a similar weighting of source-oriented and product-oriented generalizations in rating and production. Nonetheless, Figure 5 also shows that
The cluster of mappings in which some source is mapped onto [tʃi] is further subdivided into a cluster of mappings presented during training and a cluster of mappings that were not presented. This suggests that at the very least the subjects must know which source consonants need to be retained in the product form, a type of source-oriented knowledge describable using faithfulness constraints in Optimality Theory (e.g., Downing et al. 2005).

The clustering analysis in Figure 5 showed us that overall typical characteristics of products are more salient than typical characteristics of source-product mappings, even following source-oriented training, which maximizes the salience of source-product mappings. Thus tʃ→tʃi is classified as primarily X→tʃi rather than 0→i/C_. Figures 6-7 take a closer look at the data in order to explain what leads to this classification, showing that, while the participants were not exposed to alveolar consonants undergoing palatalization, velar palatalization was often overgeneralized to alveolar sources, especially following product-oriented training (the difference between training paradigms in overall resulting rate of alveolar palatalization is significant at p<.00001 according to the Wilcoxon test).

When the data from both types of training are combined, examples of tʃ→tʃi significantly favor alveolar palatalization (t→tʃi) (F(1,78)=7.7, p=.006 for production, shown in Figure 6; F(1,78)=10.9, p=.001 for rating, shown in Figure 7), and there is no significant interaction between training paradigm and whether or not examples of [tʃ]→[tʃi] are presented (F(1,78)<1, p=.77 for production; F(1,78)<1, p=.83 for rating). This relatively strong effect of exposure to tʃ→tʃi on the productivity of t→tʃi relative to t→ti is the main reason for the clustering algorithm classifying tʃ→tʃi as an instance of X→tʃi following source-oriented training.
Figure 6: Following either type of training, examples of $tʃ \rightarrow tʃi$ favor $t \rightarrow tʃi$ over $t \rightarrow ti$ in production (notches not shown since they go outside of the box).

![Box plot showing probability of merger of alveolar to alveopalatal.aining](image)

**Figure 7:** Following either type of training, examples of $tʃ \rightarrow tʃi$ favor $t \rightarrow tʃi$ over $t \rightarrow ti$ in rating.

![Box plot showing rating of $t \chi$ minus rating of $tʃ\chi$.](image)

Figure 8 shows that the addition of examples of $tʃ \rightarrow tʃi$ to training has different effects on the productivity of velar palatalization in the two training paradigms. In the source-oriented
training paradigm, examples of $t\rightarrow t\acute{i}$ support $k\rightarrow ki$ over $k\rightarrow t\acute{j}i$. In the product-oriented training paradigm, examples of $t\rightarrow t\acute{j}i$ support $k\rightarrow t\acute{j}i$ over $k\rightarrow ki$. If we combine the results from both training paradigms (entering training paradigm, whether $\acute{-i}$ is attached to $[t\acute{j}]$ in training, whether $\acute{-i}$ often attaches to $[p]$ and $[t]$ in training, and all interactions into a Friedman test (i.e., an ANOVA with a rank-transformed dependent variable) as predictors of production probability) the only significant effect is an interaction between experiment and whether or not examples of $[t\acute{j}]$ being mapped onto $[t\acute{j}i]$ are presented to the learner ($F(1,79) = 6.25, p = .01$). If we take the probability of $[k]$ being mapped onto $[ki]$ as the dependent variable, there is also a significant interaction in the same direction: the additional examples of $[t\rightarrow t\acute{j}i]$ presented during training increase the probability of eliciting the production of $[k]\rightarrow [ki]$ in the source-oriented training paradigm while decreasing the probability of eliciting $[k]\rightarrow [ki]$ in the product-oriented training paradigm ($F(1,79) = 4.02, p < .05$). There are no significant effects in the ratings task.

These results suggest that the $[t\rightarrow t\acute{j}i]$ examples support ‘just add –i’ over ‘plurals must end in [t\acute{j}i]’ in source-oriented training while the opposite is true for the source-oriented training paradigm. Thus, the characteristics that distinguish the two training paradigms are able to jointly influence how much the language learner relies on product-oriented vs. source-oriented generalizations in deriving new wordforms, thus extending the lexicon of the language. Interestingly, in the source-oriented paradigm, examples of $t\rightarrow t\acute{j}i$ support $k\rightarrow ki$ over $k\rightarrow t\acute{j}i$ while supporting $t\rightarrow t\acute{j}i$ over $t\rightarrow ti$. We will return to this apparent contradiction in the General Discussion.

Despite $t\rightarrow t\acute{j}i$ disfavoring velar palatalization in source-oriented training, the clustering solution presented in Figure 5 classifies $t\rightarrow t\acute{j}i$ as an $X\rightarrow t\acute{j}i$ mapping, rather than a $0\rightarrow i/C_-$ mapping even following source-oriented training because the effect of adding examples of $t\rightarrow t\acute{j}i$ on the productivity of velar palatalization is weaker than its effect on alveolar palatalization (the former effect failing to reach significance within task).
Figure 8: Following product-oriented training, examples of tj→[t̃j] favor k→[t̃j] over k→ki; the opposite is true for source-oriented training (notches not shown since they go outside of the box).

Figure 9 shows that product-oriented training favors unfaithful (stem-changing) stem mappings over source-oriented training for labial as well as alveolar sources as well: the incidence of a [p] being mapped onto [(p)t̃j] is much higher after product-oriented training than after source-oriented training (p = .0003 according to the Wilcoxon test). While Figure 9 suggests a trend for the addition of examples of [tj]→[t̃j] to favor [p]→[(p)t̃j], most learners never overgeneralize palatalization to labials, showing a zero probability of [p]→[(p)t̃j] and the apparent trend is not significant (p = .6 according to the Wilcoxon).
Importantly, product-oriented training does not favor *all* stem changes. In the rating task, learners were asked to rate $t\xi\rightarrow ki$, $t\xi\rightarrow ka$, $k\rightarrow t\xi a$, and $t\rightarrow t\xi a$ mappings, which do not result in a good product. Figure 10 shows that the ratings of these mappings are not improved in product-oriented training relative to source-oriented training. In order to test this observation, I entered type of product, type of source, type of training, and language into a single Friedman test. There was a significant interaction between type of product and type of training ($F(1,331)=8.58$, $p=.003$) and a significant interaction between type of source and type of training ($F(1,331)=5.45$, $p=.02$) with no significant three-way interaction between source, product, and training types ($F(1,331)=1.02$, $p=.33$). Thus, it appears that product-oriented training increases the productivity/acceptability of stem changes only when those stem changes result in a good product and especially when the source-product mapping is not presented during training.
Figure 10: The effect of training type (SO = “source-oriented”, PO = “product-oriented”) on ratings of stem-changing mappings resulting in the good product (tʃi) and bad products (tʃa, ki, ka).

While typical characteristics of product forms appear to be more salient than typical characteristics of source-product mappings, the learners’ behavior is not completely product-oriented even following product-oriented training, as has already been suggested by the finding that the cluster of mappings resulting in [tʃi] is further subdivided into the observed and unobserved mappings. First, overgeneralization of velar palatalization to labial sources following either kind of training is much less likely than overgeneralization to alveolar sources (p < .00001 for product-oriented training, p = .0002 for source-oriented training, according to the Wilcoxon test, cf. Figures 6 vs. 9). Similarly, after both types of training, k->k{i;a} mappings, which result in an unobserved product, are rated higher than tʃ->k{i;a} mappings, which result in the same unobserved product but also feature a stem change (p < .0001 after either type of training). Finally, tʃ->tʃi mappings are rated higher than k->tʃi or t->ti mappings after either type of training (p < .001) despite resulting in the same product. In the case of t->tʃi vs. tʃ->tʃi, this result holds even in Languages 1 and 2 where examples of tʃ->tʃi are never presented (p = .0001 for t->tʃi vs. tʃ->tʃi, p = .05 for k->tʃi vs. tʃ->tʃi following source-oriented training; p = .007 for t->tʃi vs. tʃ->tʃi, p = .11 for k->tʃi vs. tʃ->tʃi after product-oriented training). Thus even following product-oriented training most learners’ disprefer stem changes and possess grammars that contain source-oriented generalizations, perhaps, in the form of paradigm uniformity constraints (Becker & Fainleib 2009, Downing et al. 2005, Stemberger &
Bernhardt 1999), that allow them to restrict the types of sources that can give rise to a good product, thus avoiding mapping \([p]\) onto \((p)[t\bar{\jmath}]\).

4. Discussion

4.1. The influence of the learning situation

In the present experiments, the learners were exposed to miniature artificial languages in two different training paradigms. The two training paradigms differ in that:

1) The learner in product-oriented training is presented with one wordform from a paradigm at a time, while the learner in source-oriented training is presented with pairs of words that share the stem.

2) The learner in product-oriented training is exposed to a much smaller number of word types and a much larger number of word tokens per type than the learner in source-oriented training; one consequence of this difference is that the learner in product-oriented training can, to a large extent (empirically 92\% on average), memorize the lexicon exemplifying the grammar, while the learner in source-oriented training acquires a grammar without a lexicon.\(^3\)

When asked to go beyond the acquired lexicon and apply the learned grammar to new words, learners exposed to product-oriented training are found to exhibit stronger reliance on product-oriented generalizations and weaker reliance on source-oriented generalizations than learners exposed to source-oriented training. This manifests itself in two ways:

1) Examples of \(t\bar{\jmath}\rightarrow t\bar{\jmath}\) are taken to support \(k\rightarrow ki\), a mapping with the same source-product relationship, over \(k\rightarrow t\bar{\jmath}\), a mapping resulting in the same product, by learners exposed to source-oriented training; the opposite is true for learners exposed to product-oriented training.

2) The learner in product-oriented training overgeneralizes palatalization to alveolar and labial sources much more than does the learner in source-oriented training. That is, the product-oriented learner infers \(t\rightarrow(t)t\bar{\jmath}\) and sometimes \(p\rightarrow(p)t\bar{\jmath}\) based on exposure to \(k\rightarrow t\bar{\jmath}\) while the source-oriented learner does not.

\(^3\) In addition, in the original experiment subjects in the product-oriented paradigm signed up to ‘learn names for objects’ while the learners in source-oriented training signed up to learn ‘how to make plurals’ in a made-up language. However, I have subsequently conducted a product-oriented training experiment in which half of the subjects (N=32) were presented with each type of instruction and found no significant effect of instruction (F<1), thus the difference in instructions is unlikely to lead to the differences in behavior following the two training paradigms.
Product-oriented training does not simply influence how much the learners prefer to avoid stem changes across the board. Stem changes that do not result in a good product, e.g., \( t\rightarrow k\{a;i\} \), do not benefit from product-oriented training. Rather, product-oriented training appears to draw attention away from source-product relationships and towards characteristics of product forms (or perhaps, all individual wordforms), compared to source-oriented training.

Thus the present results support the idea that the types of generalizations that are relied upon by a speaker/hearer in extending his/her lexicon are influenced by the way the speaker/hearer experiences language, and not just by an innate Universal Grammar, suggesting that even formal properties of the grammar may be emergent from patterns of language use (Bybee 2008). As Valian and Coulson (1988: 78) suggested,

> Our … acquisition of competence is mediated by the performance system. That performance system … limits us to acquiring a language only under presentation conditions which are cognitively favorable.

The present results indicate that presentation conditions may bias a learner in favor of source-oriented or product-oriented generalizations. If native speakers of natural languages prefer product-oriented generalizations over rules (Becker and Fainleib 2009, Bybee 2001, Bybee and Slobin 1982, Köpcke 1988, Lobben 1991, Wang and Derwing 1994), this may be due to the way those languages are experienced by their native speakers, since learners tend not to hear multiple forms of the same lexeme one after another.

At least three predictions for natural languages follow from the observed effect of the learning task. First, reliance on source-oriented generalizations may be more expected in non-native speakers of a language, who experience language through textbooks that explicitly teach the reader to conjugate verbs and decline nouns, than in native speakers who experience language one wordform at a time. Second, source-oriented generalizations should form when wordforms sharing a stem tend to appear in close temporal proximity. This is, perhaps, the case for noun-adjective pairs of the type ‘electric-electricity’ in English, for which source-oriented generalizations like \( k\rightarrow s/\_\_t\) (or ‘an [l] in the noun corresponds to an [l] in the adjective’) appear to be stronger than product-oriented generalizations like ‘-ity is usually/should be preceded by [l]’ (Pierrehumbert 2006). Some support for this hypothesis is provided by Morgan, Meier and Newport (1989) who found that the acquisition of a phrase structure grammar was facilitated when learners were provided with pairs of sentences that could be
related by pronominalization or movement rules but were unable to replicate the effect with related pairs of sentences being randomly interspersed with other, unrelated sentences. Finally, product-oriented generalizations may be favored over source-oriented generalizations especially strongly if both have to be acquired over a small set of word types where the inherently lower type frequency of source-oriented generalizations may be of particular importance.

4.2. Task-independent properties of grammar

While the learning situation influences the degree to which the learner relies on source-oriented vs. product-oriented generalizations, and thus the acquired grammar, there are a number of characteristics of the acquired grammatical systems that hold across the two learning situations.

When two generalizations, or two potential product forms are in competition for the same source, the competition is resolved stochastically, as proposed by the Minimal Generalization Learner (Albright and Hayes 2003), stochastic Optimality Theory (Boersma 1997, Boersma and Hayes 2001), and noisy Harmonic Grammar (Boersma & Weenink 2009, Coetzee and Pater in press, Smolensky and Legendre 2006). The learner does not always choose to obey the most reliable generalization or produce the best-supported product, contrary to the predictions of classical Optimality Theory (Prince and Smolensky 1993 / 2004) and classical generative phonology (see also Coetzee and Pater in press). If the learner always obeyed the most reliable applicable rule, or produced the most harmonious product form, the learner would always produce a [tʃi]-final product from a [k]-final source in all artificial languages tested, since k → tʃi is more reliable than ‘just add –i’ in any language, or in an Optimality-Theoretic framework, Ident-velar is always outranked by *ki. One caveat is that the learners in the present experiments are adults. Hudson Kam and Newport (2005) show that when exposed to an unpredictable alternation between two rules, child learners tend to use the best-supported rule 100% of the time, while adults tend to match the probabilities of the rules in the input. Thus, it is important to replicate the present experiments with children.

The learners in both training paradigms learn grammars that contain both product-oriented generalizations, such as ‘plurals should end in -tʃi’, and source-oriented generalizations that restrict the sources that can be mapped onto a product (cf. also Pierrehumbert 2006). Thus, despite the overall preference for [tʃi]-final plurals, [p]-final sources are less likely to be mapped onto [tʃi]-final plurals than [t]-final or [k]-final sources even after product-oriented training (a finding that mirrors linguistic typology, as shown by Bateman 2007). In addition, stem changes resulting in unobserved products are dispreferred relative to simple addition of an
affix resulting in the same unobserved product. Thus, the learned grammar is not purely product-oriented. The product-oriented generalizations need to be supplemented with something analogous to paradigm uniformity constraints, e.g., ‘if there is a [k] in the singular, there must be a [k] (in the same position) in the plural’. The present data provide no evidence regarding whether these constraints are learned. It is quite possible that the learners come to the experiment knowing that \( p \rightarrow t \) mappings are worse than \( t \rightarrow t \) mappings. On the other hand, it is also possible that \( t \rightarrow t \) mappings are favored relative to \( t \rightarrow t \) mappings in a way that \( p \rightarrow t \) mappings are not because \( t \rightarrow t \) is acoustically more similar to \( t \rightarrow t \) than \( p \rightarrow p \) is to \( p \rightarrow t \).

The necessity of supplementing product-oriented generalizations with restrictions on which source forms can be mapped onto a desirable product (i.e., paradigm uniformity constraints, see Becker and Fainleib 2009, Downing, Hall and Raffelsiefen 2005, Stemberger and Bernhardt 1999) is also suggested by Pierrehumbert (2006). Pierrehumbert shows that when a native English speaker is presented with a novel Latinate adjective ending in [k] and produces a noun ending in –ity from it, as in ‘interponic’ \( \rightarrow \) ‘interponicity’, the adjective-final [k] is changed into an [s] when followed by –ity. Pierrehumbert argues that English speakers must be using a source-oriented generalization like k\( \rightarrow s/\_i \) and not a product-oriented one like ‘Latinate nouns should end in [s/\_i]’ or ‘Latinate nouns should not end in [k/\_i]’ for two reasons. First, only adjectives ending in [k] are mapped onto nouns ending in [s/\_i]. This shortcoming is remedied by allowing segment-specific paradigm uniformity constraints like ‘a [t] present in the adjective is retained in the noun’, which, being made over source-product pairs, are source-oriented generalizations. Second, Pierrehumbert shows that [s] is not the consonant that most commonly precedes –ity in English. Rather, [l] precedes -ity much more commonly than [s] does. Therefore, a learner generalizing over nouns would be expected to believe that –ity should be preceded by [l] much more often than by [s]. Nonetheless, speakers in Pierrehumbert’s experiment never changed [k] into [l] when attaching –ity. Generalization over adjective-noun pairs, on the other hand, would yield the observed pattern of [k] being mapped onto [s] and not [l] because adjectives ending in [k] never correspond to nouns ending in [l/\_i] but often correspond to nouns ending in [s/\_i].

Generalization is not minimal in the present study. This is a violation of the popular Subset Principle (Berwick 1986, Dell 1981, Hale and Reiss 2003). It appears worthwhile to distinguish between two types of overgeneralization. One type of overgeneralization is, I would argue, an inevitable result of perceptual processes. Traditionally, the output of human perception is taken to be a single hypothesis about the identity of the stimulus, thus the only information provided
by perception is the identity of the most probable stimulus given the evidence. For instance, Clayards et al. (2008: 804), in a paper arguing for an otherwise Bayesian approach to speech perception, write “the goal of speech perception can be characterized as finding the most likely intended message”. Under a purely Bayesian approach, on the other hand, the output of perception is a probability distribution over possible stimuli (Kruschke 2008, Levy 2009). Thus, despite reporting having perceived the most probable stimulus, the perceiver assigns other similar stimuli non-zero probabilities of having been presented. For instance, a subject presented with [ti] may report hearing [ti] but also (subconsciously) consider it possible but less likely that [ki] has just been presented. Note that if the learner intends to maximize the probability of being correct, s/he should always report hearing the stimulus s/he considers to be the most probable one (Norris and McQueen 2008) but should update the probability of each possible hypothesis in proportion to how likely s/he believes it to be given the sensory data (Kruschke 2008, Levy 2009).

Given these assumptions, it appears unsurprising that palatalization is much more likely to be overgeneralized to [t] than to [p] and that palatalization is overgeneralized to [t] despite accurate reporting of hearing t\rightarrow ti when presented with t\rightarrow ti. It appears inevitable that a perceiver hearing (and reporting hearing) [t^{(0)}i] would assign some probability to having heard [t^{(j)}i] and that this estimated probability would be higher when [t^{(0)}i] is presented than when [p^{(0)}i] is presented. Thus, overgeneralization of palatalization to [t] is predicted to be more likely (perhaps, inevitable) given Bayesian perception, than overgeneralization to [p], which appears to be ‘genuine’ overgeneralization due solely to the product-oriented schema ‘plurals must end in -ti’.

While the learned grammar contains both product-oriented and source-oriented generalizations, learners appear to pay less attention to the source-product relationship than to the shapes of typical products in both training paradigms. Thus, even after source-oriented training, the mapping t^{(j)}\rightarrow t^{(j)}i is treated as more similar to other mappings resulting in the same product (t^{(j)}i) than to other mappings featuring the same source-product relationship ([ ]\rightarrow i). This finding contradicts the assumptions of rule-based models (Chomsky and Halle 1968, Albright and Hayes 2003, Plag 2003) and provides support for the product-oriented Network Theory (Bybee 2001).

An important remaining question is whether the product-oriented generalizations are positive, as in Bybee’s Network Theory (Bybee 1985, 2001) and Stemberger and Bernhardt’s version of
Optimality Theory (Stemberger and Bernhardt 1999) or negative, as in traditional (Prince and Smolensky 1993/2004) and Stochastic Optimality Theory (Boersma 1997, Boersma and Hayes 2001). Interestingly, simulations using the implementation of Stochastic Optimality Theory in Praat (Boersma and Weenink 2009) show that, despite incorporating product-oriented markedness constraints, Stochastic Optimality Theory has problems handling the present data.

The fact that learners in the present experiment appear to learn that velars and possibly alveolars become alveopalatals before –i can be modeled by the constraint weighting in (1).

Palatalization of a consonant with a certain place of articulation is triggered by the applicable *i constraint being ranked above the applicable Ident-Place constraint.

(1) *ki, Ident-Labial > > *C\textsubscript{Stop}, Ident-Alveolar, Ident-Velar, *a

Examples of t\textsubscript{f}→t\textsubscript{fi} do not provide evidence on whether Ident-Velar, Ident-Labial, and Ident-Alveolar constraints should be ranked above or below *C\textsubscript{Stop} or *ki. Thus, examples of t\textsubscript{f}→t\textsubscript{fi} should have no effect on the estimated desirability of [t\textsubscript{fi}]-final plurals relative to [ki]-final singulars relative to [ki]-final plurals. Furthermore, examples of t\textsubscript{f}→t\textsubscript{ji} provide evidence against *i > > *a, thus voting for the attachment of –i with or without a change to the preceding consonant. In combination, providing no evidence regarding the ranking of *C\textsubscript{Stop} and Ident-place and providing evidence for *i > > *a leads to increasing the probability of adding –i to a velar-, alveolar- or labial-final source without changing the consonant.

The only way for the examples of t\textsubscript{f}→t\textsubscript{fi} to, e.g., favor t→t\textsubscript{ji} over t→ti, there must be a *t\textsubscript{fi} constraint whose weight is decreased by examples of t\textsubscript{f}→t\textsubscript{fi}. Why learners should come to the task with such a constraint (which should be relatively highly-ranked for its demotion to have appreciable effects on behavior) remains a mystery since it is supported neither by training data nor the learners’ prior linguistic experience. On the other hand, in Network Theory, [t\textsubscript{fi}]-final plurals support other [t\textsubscript{fi}]-final plurals, whatever the source, because of a generalization like ‘plurals must end in -t\textsubscript{fi}’, which is supported by the training data, in which t\textsubscript{fi}-final plurals form a large proportion of the lexicon (cf. also Stemberger and Bernhardt 1999: 437-438).

An alternative way to weight a constraint against the unobserved sequence [ki] is to calculate the likelihood that the absence of [ki] is not accidental by taking the difference between how often [ki] is expected to occur and how often it actually occurs based on the frequencies of occurrence of related sequences in plural forms (Frisch, Broe and Pierrehumbert 2004, Pierrehumbert 1993, Stefanowitsch 2008, Xu and Tenenbaum 2007). The actual frequency of
occurrence of \([ki]\) is zero across the two artificial languages. However, other \([Ci]\) sequences occur much more often when examples of \(tf \rightarrow t\bar{fi}\) are presented. Thus, the learner estimating how often \([ki]\) would occur if it were just like the other \([Ci]\) sequences would estimate a higher frequency when exposed to examples of \(tf \rightarrow t\bar{fi}\), which would cause him/her to be more confident that \([ki]\) is to be avoided. For example, Xu and Tenenbaum (2007) find that learners presented with three examples of the novel word \(fep\) infer that \(fep\) means ‘Dalmatian’ rather than ‘any dog’ more often than if only one \(fep\)-Dalmatian pairing is presented. Xu and Tenenbaum argue that the learners detect a suspicious correlation between \(fep\) and pictures of Dalmatians, which would be unexpected if \(fep\) could refer to any dog. Regier and Gahl (2004) and Stefanowitsch (2008:518) propose that the same mechanism may be used in syntax. If phonology learning worked the same way (as suggested by Frisch, Broe and Pierrehumbert, 2004 and Pierrehumbert 1993 for OCP), we would expect that exposure to examples of \(tf \rightarrow t\bar{fi}\) would restrain \(-i\) from simply attaching to \([k]\). Thus, the examples of \(tf \rightarrow t\bar{fi}\) would disfavor palatalization, contrary to the data presented here as well as the data in Kapatsinski (in press), which shows that additional examples of \(p;\bar{t} \rightarrow p;\bar{t}\bar{i}\) strongly favor \(k \rightarrow ki\) rather than restricting attachment of \(-i\) to labial-final and alveolar-final sources. Thus, the present data support reliance on positive, rather than negative product-oriented generalizations (Bybee 1985, 2001, Stemberger and Bernhardt 1999).

It may be expected that constraints against unobserved combinations of units should be less salient in phonology than in lexical semantics (Xu & Tenenbaum 2007) or syntax (Regier and Gahl 2004, Stefanowitsch 2008) because unobserved unit combinations are usually more similar acoustically to observed combinations in phonology than in syntax or the lexicon. A learner hearing \([pa]\) is expected to assign some probability to having heard \([ka]\), and a learner hearing \([t^{\ddagger}i]\) or \([p^{\ddagger}i]\) may assign some probability to having heard \([k^{\ddagger}i]\) even if the correct phoneme sequence is reported. Thus exposure to phoneme sequences that are similar to an unobserved phoneme sequence should not decrease the estimated probability of the unobserved sequence if the similar sequences are similar enough to be confusable with the unobserved sequence (although the observed sequence should benefit from its presentation more than other similar sequences). Perceptual similarity between words, animal pictures (Xu & Tenenbaum 2007), or word sequences (Stefanowitsch 2008) is generally lower than between the level of phoneme sequences, thus an unobserved combination is less likely to benefit from the presentation of a similar combination. Thus, estimation of the reality of a gap based on the frequency of occurrence of related sequences may play a larger role in syntax and word learning than in phonology.
In both training paradigms, examples of t\textsuperscript{ʃi}→t\textsuperscript{ʃi} support t→t\textsuperscript{ʃi} over t→t\textsuperscript{i} and p→t\textsuperscript{ʃi} over p→p\textsuperscript{i}. In the source-oriented paradigm, the same examples also support k→k\textsuperscript{i} over k→t\textsuperscript{ʃi}. In the product-oriented paradigm, they support k→t\textsuperscript{ʃi} over k→k\textsuperscript{i} but not as much as they support t→t\textsuperscript{ʃi} over t→t\textsuperscript{i}. One thing that distinguishes t→t\textsuperscript{ʃi}, p→t\textsuperscript{ʃi}, and k→k\textsuperscript{i} from t→t\textsuperscript{i}, p→p\textsuperscript{i}, and k→t\textsuperscript{ʃi} is that the former set of mappings is unobserved during training while the latter is observed. Thus, we may hypothesize that the same amount of extra support increases the strength of a poorly supported mapping (e.g., k→k\textsuperscript{i}) more than it increases the strength of a mapping that is already well supported (e.g., t→t\textsuperscript{i}). That is, the relationship between amount of support from the training data and resulting strength of a source-product mapping or a candidate product form is a decelerating function, like a logarithm (cf. Goldiamond and Hawkins 1958 for the same effect in word recognition; Norris and McQueen 2008 for computational evidence that the decelerating function emerges out of Bayesian inference). An alternative explanation is that source-product mappings involving similar segments support each other and the learners consider [t\textsuperscript{ʃ}] to be more similar to [k] than to [t], thus t\textsuperscript{ʃ}→t\textsuperscript{ʃi} examples provide more support for k→k\textsuperscript{i} than to t→t\textsuperscript{i} and following source-oriented training the increase in support for k→k\textsuperscript{i} happens to be greater than the increase in support for ‘plurals end in [t\textsuperscript{ʃi}]’ but the increase in support for t→t\textsuperscript{i} is not.

In general, the results from elicited production and rating tasks are extremely similar. The one difference between elicited production and rating observed in the present data is that elicited production appears to disfavor stem changes more than rating does (see also Zuraw 2000 for the same finding in natural language). Thus, only 4/44 learners exposed to source-oriented training produce more instances of t→t\textsuperscript{ʃi} than t→t\textsuperscript{i} but the median difference in standardized ratings between the two mappings is only .13 (standard deviations), and 16/44 learners assign lower ratings to t→t\textsuperscript{i} than to t→t\textsuperscript{ʃi}. The median difference in production probability between k→k\textsuperscript{i} and k→t\textsuperscript{ʃi} is 0, while k→k\textsuperscript{i} is rated as being somewhat less probable than k→t\textsuperscript{ʃi} (.3 standard deviations). Nonetheless, the difference is small and significant only for the velars (p = .01, according to the Wilcoxon).

5. Conclusion
The results provide support for a grammar that 1) contains both positive product-oriented generalizations (a.k.a. schemas, Bybee 1985, 2001) and source-oriented paradigm uniformity constraints, a combination proposed by Stemberger and Bernhardt (1999), and 2) resolves competition for a source form between generalizations or potential product forms
stochastically, depending on relative acceptability of competitors, which is a decelerating function of statistical support for the competitors in the presented training data. The learner acquiring the grammar appears to 1) pay more attention to characteristics of the product than to the source-product relationship, especially when sources and products do not occur in close temporal proximity and/or the size of the lexicon exemplifying the grammar is relatively small, and 2) assign some probability mass to percepts other than the most probable one, i.e., the one the learner reports hearing.

References:
Bybee, Joan L., & Carol Lynn Moder 1983 Morphological classes as natural categories. 
*Language* 59: 251-270.


Caballero, Gabriela  In press Scope, phonology and templates in an agglutinating language: Choguita Rarámuri (Tarahumara) variable suffix ordering. *Morphology*.


Kapatsinski, Vsevolod  In press Velar palatalization in Russian and artificial grammar: Constraints on models of morphophonology. *Laboratory Phonology*. 


