SIBILANT HARMONY: INVESTIGATING THE CONNECTION BETWEEN TYPOLOGY AND LEARNABILITY

Loredana Andreea Kosa
University of Toronto

1. Introduction

Typological patterns have often been used to establish what constraints are part of Universal Grammar (Chomsky, 1981, Greenberg, 1974, Greenberg, 1978, Prince and Smolensky, 1993) and therefore readily available to children. If this connection between typology and UG is correct, rules based on the dominant pattern found in typology should be more easily acquired. In phonology, for example, Hall & Hamann(2006) investigate universal properties of assibilation based on typology and argue that these properties must be realized through constraints or constraint rankings in Optimality Theory.

Another approach to the problem, however, is to focus on the learnability of rules and argue that typology reflects a learning bias against certain rules (rare or unattested) or a learning bias for certain rules (dominant). Some of the phonological research in the last few years has focused on that exact issue trying to determine which phonological rules are learnable. Some research into that question uses Artificial Grammar experiments (Moreton, 2008, Peperkamp et al., 2006, Wilson, 2003), while other research focuses on computational modeling/simulations of acquisition to establish what is learnable (Boersma and Hayes, 2001, Hayes and Wilson, 2008).

The experiments described in this paper aim to establish if there is any connection between typological patterns and the learnability of sibilant harmony types. The experiments were designed in such a way as to also provide a comparison between different methodologies. The effectiveness of Artificial Grammar (AG) methodologies is compared by using two different experiments based on the AG paradigm to answer the same learnability question. The methodologies employed were ones previously used to answer questions in a similar vein as the current research (Moreton, 2008, Wilson, 2003).

Sibilant harmony is particularly informative because the attested cases of sibilant harmony show a disproportion of one type versus another; namely preference for [-ant] segments as triggers and [+ant] segments as targets. Sibilant harmony is the assimilation of features such as place ([+/-ant]) between sibilant ([+strident]) consonants (e.g. fricatives /s/, /ʃ/; affricates /ts/, /tʃ/). For example, /naʃ-sis-/ ('cause to like') in Aari becomes [naʃ-ʃiʃ-] because the segment /s/ harmonizes to /ʃ/ to agree with the /ʃ/ segment in the stem /naʃ/.

If it can be shown that the predominant pattern in the typology of sibilant harmony is easier to learn for speakers with no background in languages with

Actes du congrès annuel de l'Association canadienne de linguistique 2010.
Proceedings of the 2010 annual conference of the Canadian Linguistic Association.

such harmony, then this provides some basis for connecting typological patterns to learnability and, ultimately, Universal Grammar. As well, comparing results from different methodologies investigating the same question can provide insight into the appropriateness of these methodologies for learnability questions.

The paper starts off with a discussion of sibilant harmony (section 2), followed by the key research questions and hypotheses (section 3). Section 4 details methodology including participants (section 4.1), materials (section 4.2) and procedure (section 4.3). The results and a discussion follow in section 5, while section 6 concludes the paper.

2. Sibilant Harmony

Sibilant harmony refers to harmony which affects only sibilant ([+strident]) segments such as fricatives /s/ and /ʃ/, and affricates /ts/, and /tʃ/. Sibilant harmony involves assimilation of features like place, voicing, manner at a distance. It can manifest as either morpheme structure constraints or morpheme alternations. Morpheme structure constraints dictate that well-formed stems can only contain sibilants which harmonize, while morpheme alternations involve changes in affixes to harmonize with stems, like in Aari (1) where the suffix /-sis/ is realized as [-ʃiʃ] following any palato-alveolar sibilants in the stem (Rose and Walker, 2004).

(1) Aari sibilant harmony: affix alternations (Rose and Walker, 2004)
  a. [duuk-sis-] 'cause to bury'
  b. [naʃ-ʃiʃ-] 'cause to like'

A distinction is usually made between triggers and targets, particularly in the case of morpheme alternations. Triggers are the segments which drive the harmony process, while targets are the segments that harmonize in order to agree in some feature with the triggers. For example, in Aari the triggers are the [-ant] sibilants /ʃ/ and /tʃ/ in the stem and the target is the [+ant] sibilant /s/ in the affix /sis/ (as seen in 1b).

Languages with sibilant harmony can be divided into three types: symmetric harmony where triggers can be either [+ant] or [-ant] segments (2), asymmetric harmony where the triggers are only [-ant] segments (3) and asymmetric harmony where the triggers are only [+ant] segments (4).

As seen in example 2, in Ineseno Chumash both /s/ (2a) and /ʃ/ (2b) can be triggers. In (2a) the rightmost [+ant] segment triggers harmony in the preceding affricate, while in (2b) the rightmost [-ant] segment triggers harmony in two preceding fricatives.

(2) Symmetric sibilant harmony: Ineseño Chumash [data from Hansson(2001) citing Applegate(1972)]
  a. Right-to-left: /s/ [+ant] trigger:

/s-api-tʃ^{h}o-us/ → [sapits^{h}olus]
'he has a stroke of good luck'

b. Right-to-left: /ʃ/ [-ant] trigger:
/s-api-tʃ^{h}o-us-waʃ/ → [ʃapitʃ^{h}oluʃwaʃ]
'he had a stroke of good luck'

As seen in example 3, in Sarcee only the rightmost [-ant] segment acts as a trigger. In 'my duck' (3b), the rightmost /z/ [+ant] segment does not trigger harmony, while the preceding /tʃ/ [-ant] segment does cause /s/ to harmonize to /ʃ/.

(3) Asymmetric sibilant harmony with [-ant] triggers: Sarcee [data from Hansson(2001) citing Cook(1979, 1984)]

a. Right-to-left: /tʃ/ [-ant] trigger:
/si-tʃogo/ → [ʃì-tʃógò] 'my flank'

b. Right-to-left: /z/ [+ant] not a trigger, but /tʃ/ [-ant] is a trigger:
/si-tʃiz-aʔ/ → [ʃí-tʃídz-àʔ] 'my duck'

As seen in example 4, in Tepehua only the rightmost [+ant] segment acts as a trigger. In (4a) the rightmost [+ant] segment (/s/) triggers harmony in preceding sibilants changing /tʃ'/ into /ts'/ ('toe nail'). However, when the rightmost segment is a [-ant] sibilant no harmony is triggered (4b).

(4) Asymmetric sibilant harmony with [+ant] triggers: Tlachichilco Tepehua [data from Hansson(2001) citing Watters(1988)]

a. Right-to-left: /s/ and /ts/ [+ant] triggers:
/tʃ'an-q'isi:ti/ → [ts'anʔesi:ti] 'toe nail'

b. Right-to-left: /ʃ/ and /tʃ/ [-ant] are not triggers:
/tasa-ʃka-/ → [tasaʃka] 'toothache'

From all the known cases of sibilant harmony in world languages, there are 13 symmetric languages and 14 asymmetric languages with [-ant] triggers. However, there is only one asymmetric language with [+ant] triggers, Tepehua(Hansson, 2001, Kochetov et al., 2008).

Correspondence approaches to consonant harmony, such as Rose & Walker(2004) propose constraints which establish a relationship called correspondence between similar segments. Along with an agreement constraint, these constraints ensure that agreement at a distance only applies to similar segments. Faithfulness constraints complete the picture ensuring that only minimal changes from input to output are allowed. This type of approach allows for any direction of change in harmony systems, depending on the relative ranking of faithfulness constraints. In the case of sibilant harmony, with faithfulness to [+ant] marked higher, the language allows [-ant] segments to change to [+ant] segments but not the reverse, just like in Tepehua (the rare language). With faithfulness to [-ant] marked higher, the language allows [+ant] segments to change to [-ant] segments but not the reverse, just like in the majority of languages with asymmetric sibilant harmony. Because of this, correspondence theory does not make any predictions about a difference between [+ant] or [-ant] triggers, instead it allows for all three types of languages.

An alternative approach is to claim a learning bias for [-ant] triggers. Instead of a fixed ranking, it is possible for children to start with faithfulness to [-ant] ranked above faithfulness to [+ant]. This would lead to children having more difficulty learning sibilant harmony with [+ant] triggers as they must overcome an initial ranking that does not allow such rules. Such a model would account both for the learning bias and the typology of sibilant harmony, while maintaining the connection between the two.

Hansson(2001) proposes that harmony is a result of "phonologized speech errors" and therefore error patterns should be reflected in the typological tendencies of consonant harmony . In speech errors with sibilants, alveolars like /s/ are far more susceptible to interference from more palatal segments such as /ʃ/ than vice versa (Hansson, 2001). This means that alveolars are more likely to act as targets which harmonize than as triggers. This is reflected in the typological pattern of sibilant harmony where [-ant] segments are the triggers in all but one case (Tepehua).

The question still remains if difficulties with phonological planning or phonetic implementation of similar consonants are the sole reason for the typology of sibilant harmony or if there is a more general learning bias for certain types of triggers. A correlation between learnability and typology does not dismiss the importance of phonological and phonetic factors on the emergence of sibilant harmony typology. Such a result is consistent with a learning bias for palato-alveolar triggers, but also with an underlying common factor between phonological errors and learnability that influences typological patterns.

3. Research Questions

The current research focuses on the connection between typology and learnability; in particular on whether or not a predominant rule is easier to acquire compared to a rare rule. The phonological rule investigated is sibilant harmony with the predominant rule being asymmetric harmony with [-ant] triggers and the rare rule being asymmetric harmony with [+ant] triggers. Previous psycholinguistic research suggests that dominant rules are learnable, while unattested, rare or arbitrary patterns are not. As such, the hypothesis is that sibilant harmony with [-ant] triggers will be easier to learn than sibilant harmony with [+ant] triggers. This can manifest in two ways; either both rules are learnable but the predominant rule is learnt better or faster, or only the predominant rule is learnable. A second research question addresses methodology by asking if methodologies based on the same AG paradigm but differing in tasks can yield similar results. The hypothesis is that the two methodologies will yield comparable results.

4. Methodology

The focus of the experiments was the learnability of two directions of assimilation in sibilant harmony: s→ʃ, which is [-ant] assimilation (trigger [-ant] segments, target [+ant] segments) and ʃ→s, which is [+ant] assimilation (trigger [+ant] segments, target [-ant] segments). To reflect this focus, there were two distinct conditions reflected in two distinct languages. Language 1 contained forms that adhere to the [-ant] harmony rule, while language 2 contained forms that adhere to the [+ant] harmony rule. Each participant belonged to a language group and had to perform two key experiments within a sitting. The order was counterbalanced by having half of the participants perform experiment A first and the other half perform experiment B first.

Each participant performed experiment A, a perception experiment and experiment B and filled out a questionnaire midway through the sitting to establish language background. The data collected consisted of YES/NO responses for experiment A and the perception experiment and choice responses (1st or 2nd) for experiment B. Experiment A was based on the experimental design in Wilson(2003), while experiment B was based on the experimental design in Moreton(2008). Both experiments were modified to ensure that the same stimuli were used.

4.1 Participants

The participants were 24 (8 male, 16 female) self identified monolingual native speakers of English born in various countries. None of the languages spoken by the participants have sibilant harmony as a phonological rule. The participants were students at University of Toronto between the ages of 18 and 27 years old (average 22). All participants were of normal or corrected-to-normal vision and normal hearing and all background information about language experience was self-reported in a questionnaire. The participants were divided in four groups of six participants each based on (language x order). There were 12 participants for each experimental language, and within each experimental language group they were further divided in two groups per language based on which experiment was performed first.

4.2 Materials

The stimuli used in both experiment A and B, consisted of 33 words during the exposure phase and 110 words during the testing phase. The words were based on CVCV type stems with one of the three suffixes /næ/, /sʌt/ ([sʌt] or [ʃʌt]) and /ʃo/ ([ʃo] or [so]). The words used in the exposure phase were based on 11 stems with the grammatical form of each suffix, while the words used in the testing phase were based on 22 stems, 11 stems identical with the exposure stems and 11 new stems. During testing, the stems were attached to both grammatical and ungrammatical forms of each suffix. This resulted in 22 words with the /næ/

suffix, 44 words with the grammatical forms of the sibilant suffixes and 44 words with the ungrammatical forms of the sibilant suffixes.

There were three types of stems used in the experiments; non-sibilant stems such as /tʌro/ which contained no sibilants (3 out of 11), [+ant] stems such as /bætsæ/ which contained a [+ant] sibilant in the second syllable (4 out of 11) and [-ant] stems such as /motʃo/ which contained a [-ant] sibilant in the second syllable (4 out of 11). Table 1 shows examples of grammatical stimuli presented during exposure. The grammatical form of the sibilant suffixes changed depending on which language the participant was assigned to.

	Language 1:[+ant]→[-ant] s→ʃ			Language 2: [-ant]→[+ant] ʃ→s		
Suffix	næ	sʌt	ʃo	næ	sʌt	ʃo
Stem:						
tʌro	tʌro**næ**	tʌro**sʌt**	tʌro**ʃo**	tʌro**næ**	tʌro**sʌt**	tʌro**ʃo**
bætsæ	bætsæ**næ**	bætsæ**sʌt**	bætsæ**ʃo**	bætsæ**næ**	bætsæ**sʌt**	bætsæ**so**
motʃo	motʃo**næ**	motʃo**ʃʌt**	motʃo**ʃo**	motʃo**næ**	motʃo**sʌt**	motʃo**ʃo**

Table 1. Example stimuli from experiment

Experiment B called for pairs of stimuli to be presented to the participants during testing. In order to create the pairs, the stimuli were paired based on whether the stem of the word was in the exposure phase or not. One word in the pair was chosen to be ungrammatical and the other word in the pair was chosen to be grammatical using only /sʌt/ and /ʃo/ suffixes. Words with non-sibilant stems were matched with words with non-sibilant stems, while words with sibilant stems were matched with words with sibilant stems. This resulted in 12 pairs with non-sibilant stems (6 old and 6 new pairs of words) and 32 pairs with sibilant stems (16 old and 16 new pairs of words), making a total of 44 pairs of stimuli for the testing phase of experiment B. Outside of pairing grammatical with ungrammatical words and matching for type of stem, the choice of words in each pair was random. However, once randomly generated, the same list of pairs was used for all participants, the only difference being the order of presentation.

4.3 Procedure

The overall experiment consisted of three parts, experiment A, a perception experiment and experiment B. The order of experiments A and B varied with half the participants completing experiment A first and experiment B last, while the other half of the participants completed experiment B first and experiment A last. The perception experiment was run in between experiment A and B regardless of which experiment came first.

At the beginning, participants were informed of the overall structure of each experiment and then were given more detailed instructions before each experiment. Each participant filled out a consent form which laid out in broad strokes the setup of the experiments while not being very specific about the end goal. They were then setup in a sound attenuated booth with the first experiment (A or B) and given detailed instructions about each part of the experiment. During each experiment the participants were given instructions on the screen at each stage to remind them what they were supposed to do. After the first experiment, the participants were given instructions for the perception experiment and left to go through it. Once completed, each participant was asked to fill out a questionnaire asking for language background information, age, gender and birthplace. After the questionnaire, the participants were given detailed instructions regarding the second experiment (B or A) and were then allowed to complete that experiment. At the end each participant was compensated and asked to sign a receipt.

The perception experiment consisted of pairs of syllables of the type CV and CVC containing sounds used in experiments A and B. The participants were asked if the two syllables presented aurally were the same or not. The purpose of this experiment was two-fold: on one hand it ensured that the participants had no difficulty distinguishing [ts] from other segments and on the other hand it served as a distraction to minimize the interference between experiment A and B.

Experiment A consisted of an exposure phase and a testing phase. During the exposure phase the participants heard a series of 33 words presented twice in random order. The instructions were to try to recall as many of the words as possible. The words were presented one at a time with a pause in between (total 3.5 seconds) and a break after the first round of all the words. The participants chose when to start hearing the words for the second time. During the testing phase the participants were presented with one word at a time and asked if the word was one of the ones they heard during exposure; the response options were two keys, one labelled YES and one labelled NO. There were 110 words in the testing phase, which were divided in 2 groups of 55 each. The words in the testing phase were presented in random order.

Experiment B also had an exposure and a testing phase. Participants were instructed to repeat words aloud and were told that they would be tested on which words fit with the language they are hearing. The participants were told that this experiment was akin to someone telling them a new word of English and them being able to tell right away if the word could be a word of English. During the exposure phase participants heard the same 33 words as in experiment A and the words were repeated 4 times, each time in random order. During the testing phase, the participants were presented with pairs of words with a 450 ms pause in between and then asked which word in the pair fit with the language they'd been hearing in experiment B. The participants were asked to press 1 for the first word or 2 for the second word. There were 44 pairs of words all presented in one grouping and in a random order.

5. Results

5.1 Experiment A

The data collected in experiment A consists of number of YES responses for test words containing a sibilant in the stem. Several items, although included in the testing phase, were excluded from data reporting and analysis. The data with non-sibilant stems and also with the /næ/ suffix were excluded as they do not test the acquisition of harmony rule. The performance with stems that do not contain sibilants simply reflects the participants acquiring the proper underlying form of the affixes, while the performance with the /næ/ suffix reflects the participants being aware of the underlying form of the stems.

The data was collected as count data and proportions of YES out of total number of stimuli were calculated. The data was split up to account for several factors: language being used (language 1 or language 2), order of experiment presentation (experiment A first or last), whether stems were old or new (heard during exposure or not) and whether the words presented were grammatical or not.

As seen in table 2, participants chose YES more often for words based on old stems than words based on new stems and this was independent of grammaticality. Performance with grammatical old words was higher, but this is to be expected since participants were asked to identify if a word presented is one of the exposure words, which are the grammatical old words. Even the ungrammatical items have higher likelihood of recognition because the first two syllables are identical with the exposure items.

	Language 1: s→ʃ		**Language 2: ʃ→s**	
Experiment Order:	1st	2nd	1st	2nd
Old.Grammatical	77.08%	83.33%	80.21%	86.46%
Old.Ungrammatical	63.54%	70.83%	67.71%	64.58%
New.Grammatical	**57.29%**	**56.25%**	**57.29%**	**47.92%**
New.Ungrammatical	**45.83%**	**48.96%**	**56.25%**	**39.58%**

Table 2. Percent YES responses in test task (identification of memorized items) by language, order of experiment (1st or 2nd), familiarity of stem (old or new) and grammaticality of words.

The participants' performance with novel stimuli is the most telling of all the data and is therefore the main focus of the data analysis. As shown by the data in table 2, the participants chose YES more often with grammatical words for language 1, regardless of experiment order. For language 2, when experiment A occurred first, the performance with grammatical and ungrammatical words seems very similar. However, when experiment A occurred last, the participants exposed to language 2 chose YES more for grammatical than ungrammatical items. Ignoring order, performance with grammatical items is on average higher than with ungrammatical items, for both language groups.

The statistical analyses performed consisted of running repeated measures ANOVA and paired samples t-test for arcsin-stem transformations of the percentage data ($y = \sin^{-1}(\mathrm{sqrt}(x))$, where x is the original proportion) and square-stem transformations of the count data. In most cases the results from statistical analyses on both types of transformations were comparable and as such only the results on arcsin-stem transformations of percentage data are reported.

The language 1 data was entered into a 2 order (experiment A first vs. last) x 2 stem type (old vs. new) x 2 grammaticality (grammatical vs. ungrammatical) mixed subject ANOVA with the repeated measures on the last two factors. The ANOVA was run on language 1 data and showed significance for stem type ($F(1,10) = 50.261$ $p < 0.001$) and grammaticality ($F(1,10) = 14.111$ $p = 0.004$), but no significance for order ($F(1,10) = 1.201$ $p = 0.299$) or any interactions. These results mean that participants exposed to language 1 said YES significantly more to words containing old stems compared to new stems and YES significantly more to grammatical words compared to ungrammatical words, overall. However, order did not have an impact on the performance of participants exposed to language 1.

Since the new items are the ones that reflect whether or not the participants acquired the harmony rule and since the order of experiments did not show an effect, a paired samples t-test was ran on new words data for language 1 with grammaticality as the factor. There was a significant difference between performance with grammatical and ungrammatical items, $t(11) = 2.666$ $p = 0.022$. This result means that participants chose YES significantly more with grammatical new words compared to ungrammatical new words.

The same data analysis tools were used to analyse the language 2 data. The repeated measures ANOVA showed significance for stem type with $F(1,10) = 78.191$ $p < 0.001$ and grammaticality with $F(1,10) = 15.964$ $p = 0.003$, but also with the interaction (stem type * order) with $F(1,10) = 7.240$ $p = 0.023$. However, order did not reach significance overall with $F(1,10) = 3.810$ $p = 0.079$ even though the result with count data is close to significance ($F(1,10) = 4.826$ $p = 0.053$). These results mean that order affected words with old stems differently than words with new stems. With old stimuli, the average YES responses were slightly lower when experiment A occurred first than when it occurred second. However, with new stimuli, the average YES responses were higher when experiment A occurred first than when it occurred second. The interaction between stem type and order being significant means that the difference in performance between the two orders is significantly different in direction (aka sign) with old stimuli compared to new stimuli. The fact that order almost reached significance overall suggests that order may have had an impact on performance with language 2.

The new stimuli data was entered in 2 order (experiment A first vs. last) x 2 grammaticality (grammatical vs. ungrammatical) mixed subject ANOVA with the repeated measure on the last factor. The ANOVA showed no significance for the interaction between order * grammaticality ($F(1,10) = 0.486$ $p = 0.502$, only on square-stem transformation count data). This suggests that the difference between grammatical and ungrammatical stimuli is not affected by experiment order. A paired samples t-test on new words with grammaticality as the factor showed no significant effect with $t(11) = 0.867$ $p = 0.404$. A paired samples t-test on just the new words data for participants who did experiment A last, showed no significant effect of grammaticality with $t(5)=1.525$ $p=0.188$. The effect of grammaticality was higher with only the data from 2nd order, but still not high enough to reach significance.

In order to clarify the impact of order, the new stimuli data for when experiment A occurred first and the complete data for both orders were entered in a 2 language (language 1 vs. language 2) x 2 order (experiment A first vs. last) x 2 grammaticality (grammatical vs. ungrammatical) mixed subject ANOVA with the repeated measure on the last factor. No interactions reached significance and only grammaticality overall reached significance with $F(1,10) = 4.756$ $p =0.037$. This suggests that when comparing only new stimuli data from experiment A occurring first with the entire new stimuli data set for both orders, order makes no difference. As expected, grammaticality did make a difference when averaging over the entire data set, regardless of language and order. The lack of interaction between grammaticality and language suggests that the difference between grammatical and ungrammatical items did not change direction (aka sign) between the two languages. Even though the new stimuli data for language 1 showed a significant difference in performance between grammatical and ungrammatical stimuli and language 2 data did not show a significant difference, the language 2 group did not choose YES more for ungrammatical rather than grammatical stimuli.

Comparing the language data separately, each new stimuli data set was entered into a 2 order (experiment A first vs. all orders) x 2 grammaticality (grammatical vs. ungrammatical) mixed subject ANOVA with the repeated measure on the last factor. For language 1, there was no interaction and no significance for order with $F(1, 10) = 0.010$ $p = 0.923$. Grammaticality did reach significance with $F(1,10) = 9.924$ $p = 0.006$. This means that for language 1 data, order did not have an impact on the results. For language 2, there was no interaction and no significance for either grammaticality or order. This means that order did not have an impact on the results for language 2.

Running t-tests only on new stimuli data for when experiment A occurred first, resulted in no significance reached for either language. However, the results for language 1 were closer to significance than those for language 2; $t(5) = 1.840$ $p = 0.125$ versus $t(5) = 0.104$ $p = 0.922$. Using only the data from six participants was not enough to pick up the difference in performance with grammatical versus ungrammatical stimuli, but the t value for language 1 is larger than that for language 2.

The results from the statistical tests show that participants exposed to language 1 chose YES significantly more often for grammatical new stimuli than ungrammatical. On the other hand, the tests show that participants exposed to language 2 did not choose YES significantly more often for grammatical new stimuli than ungrammatical.

5.2 Experiment B

The data collected in experiment B consisted of correct responses on a forced choice task (choose word that fits language). All but the data on novel items were excluded from analysis because performance with items containing old stems may reflect memory as well as acquisition of harmony rule. Similarly, data with non-sibilant stems was excluded as it may reflect just the acquisition of the correct underlying form for the suffixes and not the acquisition of the harmony rule.

As seen in table 3, there was no consistent difference between old and new words (for the group performing experiment B for language 1 first, the performance with old items was lower with 57% versus 63% with new items). The group of participants running experiment B 2nd for language 1 were more accurate with old items (63%) than with new items (54%). The same holds true for language 2 participants, 66% for old items versus 60% for new items when experiment B occurred first and 63% for old items versus 61% for new items when experiment B occurred second. Figure 3 shows the overall average (ignoring order) for the key data on novel stimuli that contain sibilants in stems.

	Language 1: s→ʃ		**Language 2: ʃ→s**	
Experiment order:	1st	2nd	1st	2nd
Old Stems	**57.29%**	**62.50%**	**65.63%**	**62.50%**
New Stems	**62.50%**	**54.17%**	**60.42%**	**61.46%**

Table 3. Percentage correct responses in test task (identification of grammatical item) by language, order of experiment (1st or 2nd) and whether the words used contained old (from exposure, familiar) or new stems.

Similar to the experiment A data, the experiment B data was transformed using both arcsin-stem transformation on percentages and square-stem transformation on count data. All statistical tests were performed with both types of data, but since the results are comparable only the arcsin-stem transformation on percentages results are reported. The data for novel words was compared to chance performance (50%) using a one sample t-test for each language. For language 1, the performance was significantly better than chance with $t(11) = 3.043$ $p = 0.011$. Similarly, for language 2 the performance was significantly better than chance with $t(11) = 3.525$ $p = 0.005$.

Percentage and count data measuring correctness were each entered into a 2 language (language 1 vs. 2) x 2 order (experiment B first vs. last) between subjects ANOVA. The results showed no effects of language ($F(1,23) = 0.492$ $p = 0.491$) or order ($F(1,23) = 0.909$ $p = 0.352$).

The statistical tests performed on experiment B data suggest that both groups of participants performed better than chance and that there is no significant difference between participants exposed to language 1 versus those exposed to language 2 with respect to their correctness in choosing the grammatical item. Also, the order in which the experiments occurred did not have an impact on the results.

5.3 Discussion

The results of experiment A alone seem to support the first hypothesis, as the performance of participants exposed to the predominant rule reached criteria while the performance of participants exposed to the rare rule did not. This suggests that the predominant rule was learnt, while the rare rule was either not learnt at all or not as well. However, the results of experiment B do not support the first hypothesis, as both groups of participants performed comparably well. There is no support for the second hypothesis, that methodologies based on the AG paradigm should yield comparable results, as the results for experiment A show a difference between the groups and the results for experiment B show no significant difference between the groups.

The two methodologies differ in some key aspects which could have led to the difference in results. While both experiments used the same stimuli, in experiment B participants heard the same stimuli four times rather than the two times in experiment A, providing more input for learning the rule. Perhaps sibilant harmony with [+ant] triggers is easier and faster to acquire, but with enough data even sibilant harmony with [-ant] triggers can be acquired.

The testing phase of the two experiments also differed in some important ways. For one, the number of items in each experiment was different with experiment A using 110 test items and experiment B using 44 pairs of words. Secondly, the task for experiment A was to identify if a word had been heard before or not, YES or NO, while the task for experiment B was a forced choice task where participants had to choose which item fit with a language. The number of items in the testing phase affects the effectiveness of statistical tools in capturing real differences between groups. With fewer items in experiment B, perhaps the statistical tools did not have enough power to pick up an actual difference between the two groups. However, while there was no statistically significant difference, the language 2 participants performed better than language 1 participants. Therefore, it is unlikely that the statistical tools would pick up a reverse pattern with further data and as such statistical power may not be enough to explain the difference between methodologies. The difference in task types could have also had an effect on the results, however, different statistical tools and criteria were employed to suit the tasks in each experiment so the difference in task type should not be a factor.

The original experiment design in Moreton(2008) called for a lot more testing items, but, in an attempt to ensure that the exact same set of words was used in both experiments, experiment B used only 44 pairs. It is possible that the methodology employed by Moreton(2008) requires more testing items in order to capture real differences between groups.

A final key difference between the two methodologies is that only experiment B called attention to the fact that this is a learning scenario. Experiment A made no reference or mention of learning patterns, in fact it distracted participants by asking them to memorize rather than pay attention to word patterns. Experiment B informed participants that they are trying to learn a new language and must be able to pick up words that are grammatical/fit with this new language. Because of this key difference, it can be argued that experiment A more closely resembles first language acquisition, while experiment B is closer to second language acquisition. The aim of the experiments was to tap into the acquisition abilities of participants rather than their ability to actively find patterns. The difference in results could be argued to be proof that the two methodologies tap into different learning skill sets.

If the results of the methodologies are taken at face value, assuming that the difference is picking up important distinctions rather than problems in methodology and the experiments are measuring learnability, there is an interpretation which can account for the difference in results. Experiment B provided participants with twice the amount of exposure in comparison to experiment A and the results from experiment B showed no difference between the two languages while the results from experiment A showed better performance with language 1 (dominant pattern). This can be explained by an initial bias/advantage for palatalizing type harmony ([-ant] triggers) which can be overcome with sufficient exposure to fronting type harmony ([+ant] triggers).

An interpretation of this initial bias/advantage, in Optimality Theory terms, is that children start off with a default ranking which favours palatalizing harmony. Within this approach, input with fronting harmony would generate errors and with sufficient input the two faithfulness constraints would be re-ranked to allow for fronting type harmony. Because palatalizing type harmony does not require a re-ranking of constraints, it is acquired quicker and with less input compared to fronting type harmony. If this is the case, a more limited amount of input would lead to a difference between the learnability of each type of sibilant harmony, while more input would lead to no difference.

In experiment A, the amount of stimuli is more limited and therefore there is a difference in learnability between the palatalizing and the fronting harmonies. In experiment B, there is twice as much input and there is no difference in learnability between the two harmony patterns. Furthermore, when participants learning fronting harmony were exposed to experiment A last, there was an improvement in performance. Even though this improvement did not reach statistical significance, the effect of grammaticality was stronger for participants performing experiment A last rather than first. In other words, some of the extra exposure from experiment B was enough to slightly improve the performance with fronting type harmony (1% versus 8% difference) while no such improvement was observed for palatalizing type harmony. This provides further support for the idea that fronting type harmonies do benefit from more input, while palatalizing type harmonies do not.

6. Conclusion

The results of the experiments were not conclusively in support of the hypothesis that the predominant rule is easier to learn than the rare rule, but the differences between results offer strong possibilities for further insight. The results of experiment A support the hypothesis showing that the predominant rule is learnt but the rare rule is either not learnt at all or not as well. The results of experiment B, on the other hand, showed no statistically significant differences between the two rules. While not statistically significant, improvements in performance with the rare rule when more input is provided suggest that there may be a learning bias for palatalizing sibilant harmony which can be overcome with sufficient input. If taken at face value, the results of the current experiments support a correlation between learnability and typology.

The results can be used to support either an interpretation that phonology encodes this correlation as an initial bias for palatalizing sibilant harmony (or for preserving palatal sibilants over alveolar sibilants) or an interpretation that there is a learning bias within acquisition for palatalizing type harmonies which need not be encoded within phonology. The results do not dismiss the correlation between phonological errors and typology and can even be interpreted to suggest a common factor between learnability and errors which accounts for the typological patterns. The results are consistent with both an innate interpretation (within phonology or acquisition theory) and an interpretation where an underlying factor can explain both the correlation between learnability and typology and the correlation between errors and typology. The results also suggest that learning biases, at least in sibilant harmony, are of the type that aid in the quick acquisition of some processes but do not block the availability of alternative patterns. The results also suggest the significance of input in the acquisition of phonological processes.

The current findings bring to light the problems of investigating similar questions with different methodologies. While a theoretical explanation has been provided for the difference in results from the different methodologies, there is no certainty that the same skill set is assessed by different methods or that the results from one methodology are comparable to the results from another methodology. In order to provide support for learning biases, especially as they relate to typology, AG methodologies must become equivalent. When expanding psycho-linguistic methods past their original scope and design (which was proven effective), we must ensure that new versions are still equivalent in what is tested and to what degree.

As seen in this case, differences in results can lead to more questions; interesting questions, but which confuse the original endeavour. As suggested, it may be the case that the differences in results reflect a real distinction, but it can also be the case that the two methodologies measure different learning abilities. In order to test the equivalency of AG methodologies, future research can compare performance between well attested rules versus non-existent rules so as to remove the possibility that amount of input can overcome a learning bias as is, potentially, the case here.

References

Applegate, Richard B. 1972. Ineseño Chumash grammar, University of California: Doctoral dissertation.

Boersma, Paul, and Hayes, Bruce. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32:45-86.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Cook, Eung-Do. 1979. Palatalization and related rules in Sarcee. In *Linguistic studies of native Canada*, eds. Eung-Do Cook and Jonathan D. Kaye, 19-35. Vancouver: University of British Columbia Press.

Cook, Eung-Do. 1984. *A Sarcee grammar*. Vancouver: University of British Columbia Press.

Greenberg, Joseph H. 1974. *Language typology: a historical and analytic overview*. The Hague: Mouton.

Greenberg, Joseph H. ed. 1978. *Universals of human language*. Stanford: Stanford University Press.

Hall, T.A., and Hamann, Silke. 2006. Towards a typology of stop assibilation. *Linguistics* 44:1195-1236.

Hansson, Gunnar. 2001. Theoretical and typological issues in consonant harmony, University of California, Berkeley: Doctoral dissertation.

Hayes, Bruce, and Wilson, Colin. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39:379-440.

Kochetov, Alexei, Al Khatib, Sam, and Kosa, Loredana Andreea. 2008. Areal-typological constraints on consonant place harmony systems. Chicago.

Moreton, Elliott. 2008. Analytic bias and phonological typology. *Phonology* 25:83-127.

Peperkamp, Sharon, Skoruppa, Katrin, and Dupoux, Emmanuel. 2006. The role of phonetic naturalness in phonological rule acquisition. In *Proceedings of the 30th Annual Boston University Conference on Language Development*, eds. David Bamman, Tatiana Magnitskaia and Colleen Zaller, 464-475. Sommerville, MA: Cascadilla Press.

Prince, Alan S., and Smolensky, Paul. 1993. Optimality Theory: Constraint Interaction in Generative Grammar. Ms., *Rutgers University and University of Colorado, Boulder*.

Rose, Sharon, and Walker, Rachel. 2004. A typology of consonant agreement as correspondence. *Language* 80:475-532.

Watters, James. 1988. Topics in Tepehua grammar, University of California: Doctoral dissertation.

Wilson, Colin. 2003. Experimental investigation of phonological naturalness. In *Proceedings of the 22nd West Coast Conference on Formal Linguistics*, eds. G. Garding and M. Tsujimura, 533-546. Sommerville, MA: Cascadilla.