Learning a Subtractive Morphological System: Statistics and Representations

Vsevolod Kapatsinski

1. Introduction

Suppose that a company invents a new must-have gadget, called *bizwik*. Soon, everyone must have one, and these gadgets are everywhere. Even your mom has five or six shiny little _________. If you are like most English speakers, you have just generated a plural form for a noun you have never encountered before, and I'd bet that form was *bizwiks*, with an [s] (Berko, 1958). You have just solved the Paradigm Cell Filling Problem (Ackerman et al., 2009), filling a hitherto empty cell in the *bizwik*$_{SG}$~_____$_{PL}$ paradigm. How did you accomplish this feat? There are multiple competing answers to this question, corresponding to distinct views of the morphological and phonological grammar (as well as views that dispense with the notion of a grammar, which lie beyond the scope of this paper). In particular, Bybee (1985, 2001) has argued that you have used what she calls "product-oriented" schemas, form-meaning mappings linking some phonological structure – possibly underspecified – to a morphological meaning like PLURAL. In this case, a product-oriented schema would be something like [...s]$_{word}$~PLURAL. The defining characteristic of a product-oriented generalization is that it is made by generalizing over a set of forms sharing one cell in a paradigm, e.g. plural forms.

While there is evidence for product-oriented schemas (see Bybee, 2001; Kapatsinski, 2009, 2013, for reviews), a purely product-oriented grammar is insufficient to account for all productive morphological patterns: paradigm cell filling cannot involve only product-oriented schemas because the produced form can depend on other forms of the same word (Gouskova & Becker, 2013; Pierrehumbert, 2006). For example, Russian speakers know that someone who possesses only one *flarnikrapa* (Nom.Sg) must not possess multiple *flarnikrap* (Gen.Pl), whereas someone who possesses only one *flarnikrap* (Nom.Sg) must not possess multiple *flarnikrapov* (Gen.Pl). In this case, one can only decide on the Genitive Plural form if one knows whether the Nominative Singular ends in *-a* or a consonant. Since the *-a* is absent from the Genitive Plural form, generalizing over Genitive Plural forms would not allow one to discover when one should

[*] Vsevolod Kapatsinski, University of Oregon, vkapatsi@uoregon.edu. I am grateful to Paul Olejarczuk and, particularly, Zara Harmon for running the participants by tacking this study onto the ends of their various experiments, as well as to Danielle Barth for recording the stimuli.

 Proceedings of the 41st annual Boston University Conference on Language Development, ed. Maria LaMendola and Jennifer Scott, 357-372. Somerville, MA: Cascadilla Press.

produce a bare stem (*flarnikrap*) and when one should add *-ov* to it (*flarnikrapov*). Paradigmatic mappings like $C_i\#_{Nom.Sg}{\sim}C_i ov\#_{Gen.Pl}$ and $C_i a\#_{Nom.Sg}{\sim}C_i\#_{Gen.Pl}$ appear to be necessary.

Both rule-based theories of grammar and connectionist models of morphology suggest that such mappings are discovered by generalizing over pairs of morphologically related words (e.g. Albright & Hayes, 2003; Rumelhart & McClelland, 1986). Rule-based theories make the further assumption that mappings are grouped together on the basis of the *changes* they demand – the ways in which the input/source/base must be transformed into the output/product. Learners are then assumed to generalize over contexts in which the same change is observed but not over the changes themselves (Albright & Hayes, 2003; Becker & Gouskova, 2016). For example, the $C_i\#_{Nom.Sg}{\sim}C_i ov\#_{Gen.Pl}$ mapping is taken to be represented as $0 \rightarrow ov/C\_\_$ (an addition operation) while the $C_i a\#_{Nom.Sg}{\sim}C_i\#_{Gen.Pl}$ mapping is represented as $a \rightarrow 0/C\_\_$ (a subtraction operation). A rule-based learner – under this definition of 'rule-based' – will determine the contexts in which to apply each operation, but will not generalize over plural forms produced in different ways to induce a plural schema to strive for. Finally, there are also proposals that rule-like source-oriented generalizations and product-oriented schemas co-exist in the grammar, and can both influence the form that is constructed to fill an empty cell in a morphological paradigm (Nesset, 2005, 2008; Kapatsinski, under contract; Pierrehumbert, 2006).

Most patterns can be represented either as mappings between product-oriented schemas and as rules. However, the two theories of grammar can make different predictions about how learners will generalize beyond the experienced data. In this paper, I examine one such case, a subtractive morphological system, which allows us to ascertain whether – in the earliest stages of learning – the grammar is product-oriented, source-oriented or mixed. I argue for a mixed position. Namely, I propose that learners try to discover and use all possible cues to what they should produce, and that apparent preferences in favor of some cues (i.e. semantic ones, which define a product-oriented schema) emerge from these cues being easier to extract in comprehension or utilize in production.

2. The languages and predictions

In this experiment, native English speakers were presented with miniature artificial languages. In all languages, plurals were formed from singulars by subtracting the final vowel. According to rule-based theories of grammar, our participants should generalize over singular-plural word pairs to discover this vowel subtraction rule, $V \rightarrow 0/\_\_\#$. However, because all singulars presented in training were trisyllabic (CVCVCV), all plural forms had the same shape (CVCVC). A product-oriented learner who generalizes over plural forms will therefore discover an alternative generalization, the schema "plurals must have the shape CVCVC". The two generalizations make divergent predictions for shorter, CVCV singulars. The rule predicts that participants should delete the final vowels of such singulars, producing a CVC plural. The schema instead predicts that a consonant should be added to such singulars, producing a CVCVC plural like the ones that the learners have experienced.

Table 1. Training stimuli in the baseline language

Singular	Plural
baloki	balok
kiruko	kiruk
borena	boren
dalefu	dalef
farisa	faris
kalupa	kalup

Table 2: Crucial test stimuli. The training stimuli in Table 1 were also included. Test stimuli were the same across languages.

Singular	Plural
basi	?
bura	?
kali	?
kame	?
komo	?
kona	?
meni	?
taru	?

Note that the addition demanded by the schema has an inherent disadvantage in this language: subtraction results in a unique output, while addition requires the speaker to choose a consonant to add. In Kapatsinski (2013), I proposed that schemas must fully specify any additions they demand in order to be utilized. For this reason, I also presented participants with languages in which additional CVCVk stimuli were added to the training with the aim of ensuring that one particular consonant is over-represented in the stem-final position; in essence, telling participants that they should add [k] rather than some other consonant. These stimuli are shown in Table 3.

As Table 3 shows, the extra CVCVk stimuli were paired either with PLURAL meanings (PO language), SINGULAR meanings or both SINGULAR and PLURAL meanings. Recall that the product-oriented schema hypothesis predicts that schemas for forming plurals are learned by generalizing over plural forms. They are generalizations about what plurals are like. This implies that the goodness of a schema is determined either by its frequency in plurals (Bybee, 1985, 2001) or, perhaps, by the conditional probability of the schema given the plural meaning (Kapatsinski, 2013). Crucially, under this definition, the goodness of a schema is not influenced by how frequently the form specified by the schema occurs in other meanings. The form can be useless as a cue to plurality, occurring equally often in singulars and plurals, and yet be an excellent plural schema

(Kapatsinski, 2013). The prediction that follows is that the addition of final [k] to CVCV singulars should be favored when CVCVk forms are added to the plural cell of the paradigm, whether or not the same exact forms are also added to the singular cell. In contrast, adding CVCVk forms to both cells of the paradigm may be predicted to instead disfavor [k] addition if good schemas should be good cues to the meanings they express, reflecting the probability of the meaning given the form specified by the schema (e.g. Köpcke & Wecker, 2017).

Table 3 also shows one additional manipulation. While only CVCVk forms paired with plurals are added to training in the PO language, the SO+PO language also presents participants with the corresponding CVCVkV singulars. This language is intended to provide additional evidence both for the V→0 rule and for the "plurals should be CVCVk" schema relative to the baseline language, and to provide additional evidence for the V→0 rule relative to the PO language. Thus, this language is intended to test whether additional examples of CVCVkV provide more support for the rule or the schema, and whether they provide any additional support for the rule at all. Previously, Kapatsinski (2013) showed that examples of tʃ→tʃi were taken by learners to provide support for t→tʃi (same schema) over t→ti (same change). This manipulation thus serves as a conceptual replication of that study.

Table 3. Stimuli added to training in the experimental languages.

SO+PO Language		**PO Language**		**SO Language**		**Homophone Language**	
CVCVk in the right meaning		CVCVk in the right meaning		CVCVk in the wrong meaning		CVCVk in both meanings	
SG	PL	SG	PL	SG	PL	SG	PL
korika	korik	---	korik	korik	---	korik	korik
maliko	malik	---	malik	malik	---	malik	malik
meniki	menik	---	menik	menik	---	menik	menik
penuki	penuk	---	penuk	penuk	---	penuk	penuk
pineko	pinek	---	pinek	pinek	---	pinek	pinek
selaki	selak	---	selak	selak	---	selak	selak
stanoka	stanok	---	stanok	stanok	---	stanok	stanok

3. Methods

The experiment consisted of three blocks: the training/exposure block followed by an elicited production test and a judgment test. The order of blocks was fixed, so that the options presented during the judgment test would not bias production responses. Unfortunately coding of the production responses is not yet complete. For this reason, we concentrate on the judgment responses in the analyses below.

During training, all wordforms were presented auditorily in random order. They were paired with pictures of referents (depicting either one novel creature or multiple identical creatures from the videogame Spore). On each trial, the picture would appear first, followed by the onset of the word 500 ms later. The picture stayed on screen throughout the duration of the word and disappeared immediately afterwards. This order of presentation was designed to reflect the fact that caregivers tend to name objects a child is already looking at (Pereira et al., 2014). The next picture appeared 1500 ms after the word offset. Participants were instructed that they will see pictures of creatures (sometimes one, sometimes more than one), and that the names of the creatures will change depending on whether there is only one creature on the screen or many. They were told to learn the names and what singular and plural forms are like in this language. Each plural form was presented 4 times, while each singular form was presented 6 times to reinforce the singular's status as the base form (Albright, 2008; Bybee, 2001).

The production test presented participants with all the singular stimuli in Tables 1-3, paired with pictures of single creatures. The picture appeared 250 ms before the onset of the singular wordform, which was followed immediately by the picture of the plural referent. Participants were told that they will be presented with singular forms of words and that they need to say what the right plural form of each word is. They were told not to use English plurals. They had to name each picture aloud within 4 seconds. The responses were recorded. The experiment advanced to the next trial once 4 seconds elapsed.

The judgment test, on which we concentrate here, presented participants with the crucial singulars in Table 2. Each trial began with a picture of the singular referent, time-aligned with the onset of the singular wordform. 300 ms after the offset of the singular wordform, the plural wordform began, and the picture of the plural referent appeared. The picture of the plural referent was accompanied by one of two possible plurals. One possible plural involved subtraction (e.g. *bura*$_{\mathrm{SG}}$~*bur*$_{\mathrm{PL}}$). The other involved addition of [k], e.g. *bura*$_{\mathrm{SG}}$~*burak*$_{\mathrm{PL}}$. The order of judgment trials was randomized, with the aim to allow participants to accept or reject both addition and subtraction for a particular stimulus (see Kapatsinski, 2005, for discussion). On each trial, participants were asked to rate, on a scale from 1 to 5, whether they believe that the presented plural is the right plural for the presented singular, with ☹ (1) corresponding to "almost certainly NOT the right plural" and ☺ (5) corresponding to "almost certainly the right plural". They responded using a five-button serial response box with ☹ over the leftmost key, and ☺ over the rightmost one. Participants had 3 seconds to respond. The experiment advanced to the next trial upon response or if 3 seconds elapsed. At that point, there was a 1 second delay until the next trial.

All stimuli were recorded by a phonetically trained native female speaker of American English. Stress was on the initial syllable in both the singular and the plural form. With this stress location, both addition and deletion could be performed without shifting the stress even with short singulars.

40 participants were exposed to each of the five languages described above (total N=200), with each participant experiencing only one of the languages. Participants were undergraduate students in psychology or (much more rarely) linguistics recruited through the University of Oregon Psychology/Linguistics

Human Subjects Pool and received course credit for their participation. Participants were assigned to languages in the order they came into the room (e.g. first participant → baseline language, second → SO+PO Language, third → SO language etc.) Participants self-reported to be native English speakers with no history of speech, language or hearing impairment.

Data were analyzed using mixed-effects logistic regression models, as implemented in the lme4 package (version 1.1-11; Bates et al., 2015) in R 3.2 (R Core Team, 2015), fit using the "bobyqa" optimizer. The fixed-effects predictors are either Language (Tables 4 and 6) or Meaning boosted for CVCVk (Table 5) and type of judgment trial (whether the Operation it involved Addition of [k] or Subtraction of the final vowel). We concentrate on interactions between these predictors: Language or Meaning should affect how Addition is judged relative to Subtraction. Except where noted, the random effects included random intercepts for subjects and singulars, a random slope for Operation within subject and random slopes for both main effects and the interactions of interest within singular item. These constitute the maximal random-effects structure for this experimental design (Barr et al., 2013).

4. Results

Figure 1 presents the results of the judgment test. Numerically, participants exposed to the baseline language – where CVCVk was barely more frequent than other CVCVC structures – preferred subtraction in CVCV over addition of [k]. This preference for addition over subtraction reversed when additional examples of CVCVk were added, at least when this addition included added examples of CVCVk paired with the right, plural meaning – i.e., in the languages depicted in the bottom panel of Figure 1.

Tables 4 and 5 below provide data on the statistical reliability of the differences between languages seen in Figure 1. Table 4 shows that adding examples of CVCVk paired with the right meaning (plural) to training increases the judged acceptability of [k] Addition ($CVCV_{SG}$→$CVCVk_{PL}$). In contrast, adding examples of CVCVk paired with the wrong, singular, meaning has no significant effect on these judgments. In addition, subtraction is judged less acceptable after exposure to a language in which extra examples of CVCVk are added *only* to the right meaning (in the SO+PO and PO languages) than after exposure to the baseline language. Interestingly, the preference for addition holds even for the SO+PO language, where training is augmented by both additional examples of CVCVk~PL and additional examples of subtraction.

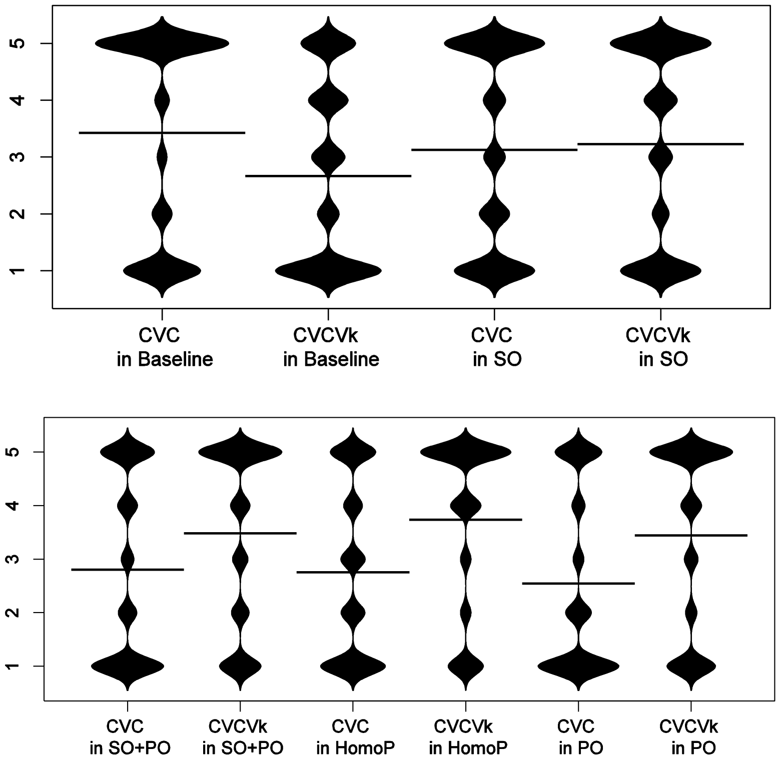

Figure 1. The effect of adding CVCVk forms paired with the wrong, Singular meaning. There are two beanstalks for each language. The left beanstalk in each pair shows ratings of subtraction (CVC), while the right shows ratings of [k] addition (CVCVk). The top panel shows the baseline language and the language in which CVCVk forms were added to the wrong singular meaning. The bottom panel shows languages in which CVCVk forms were added to the right, plural meaning, along with additional CVCVkV singulars (left), homophonous CVCVk singulars (middle), or no singulars at all (right).

Table 5 shows the results of a model that compared languages in which additional CVCVk examples were paired with the right meaning (PO and SO+PO) and the language in which they were paired with both meanings (Homophones) to the language in which they were always paired with the wrong meaning (SO). The languages in which CVCVk examples were paired with the right meaning at least half the time show reliably higher ratings of addition. When the additional CVCVk examples were *always* paired with the right meaning, ratings of subtraction were also significantly depressed.

Table 4. Report of the fixed effects from a generalized mixed effects model. Reference level for 'Language' is Baseline. Reference level for "Addition vs. Subtraction" is Subtraction.[1]

	b	*se(b)*	*z*	*p*
(Intercept)	1.70	0.44	3.85	.0001*
Language = SO	-0.23	0.48	-0.49	.62
Language = SO+PO	-1.03	0.43	-2.40	.02*
Language = Homophones	-0.62	0.49	-1.27	.20[2]
Language = PO	-1.15	0.50	-2.32	.02*
Addition	-1.25	0.52	-2.41	.02*
SO:Addition	1.00	0.72	1.39	.16
SO+PO:Addition	2.82	0.68	4.12	<.0001*
Homophones:Addition	2.53	0.76	3.32	.001*[2]
PO:Addition	2.42	0.75	3.21	.001*

Table 5. Report of the fixed effects from a generalized mixed effects model investigating whether the meaning with which additional examples of CVCVk are paired makes a difference. Reference level for Meaning is the 'wrong meaning' (SO language). Reference level for "Addition vs. Subtraction" is Subtraction.[3]

	b	*se(b)*	*z*	*p*
(Intercept)	1.43	0.39	3.69	.0002*
CVCVk boosted in both meanings	-0.40	0.43	-0.93	.35
CVCVk boosted in the right meaning	-0.78	0.35	-2.19	.03*
Addition	-0.23	0.48	-0.47	.64
Addition:BothMeanings	1.54	0.72	2.14	.03*
Addition:RightMeaning	1.44	0.57	2.55	.01*

[1] The model with the Language-by-Operation interactions is better than the model without these interactions according to the log likelihood test ($\chi^2(4)=22.14$, $p = .0002$ for the model with all five languages and no random slopes within singular; $\chi^2(3)=16.80$, $p = .0008$ for the model without the Homophones language and full random-effects structure.

[2] The model with random slopes within singulars did not converge if the Homophones language was included in the model and Language rather than Meaning was used as a predictor. For this reason, the results for this row come from a model that included Meaning rather than Language (i.e. collapsing PO and SO+PO languages). The non-converged model with Language showed the same pattern of results, with the same coefficients being significantly different from zero; see Barr et al. (2013) for dealing with non-convergence.

[3] The Boost-by-Operation interaction significantly contributes to model fit; $\chi^2(2)=7.04$, $p=.03$, according to the log likelihood test.

Table 6 shows that the languages in which additional examples of CVCVk are paired with the right meaning at least half the time do not significantly differ from each other. Thus, the Homophones language does not differ significantly from either the SO+PO language or the PO language. Neither does the SO+PO language differ from the PO language. Graphically, this lack of difference can be observed in the bottom panel of Figure 1. These results suggest that – as long as CVCVk is boosted in the plural – what happens in the singular makes little difference.

Table 6. Report of the fixed-effects from a generalized mixed effects model investigating whether what happens in the singular matters given that CVCVk is boosted in the plural. Reference level for Language is PO (i.e., no additional singular examples provided). Reference level for "Addition vs. Subtraction" is Subtraction.[4]

	b	*se(b)*	*z*	*p*
(Intercept)	0.55	0.35	1.56	.12
Language = Homophones	0.55	0.46	1.21	.23
Language = SO+PO	0.24	0.41	0.59	.55
Addition	1.08	0.46	2.38	.02*
Addition:Homophones	-0.04	0.67	-0.06	.95
Addition:SO+PO	0.02356	0.60865	0.039	.97

5. Discussion

Language comparisons in Tables 4 and 5 indicate that providing learners with additional examples of CVCVk plurals increased the ratings of $CVCV_{SG}$→$CVCVk_{PL}$ relative to the ratings of $CVCV_{SG}$→CVC_{PL}. This occurred even when the additional examples also constituted additional examples of subtraction ($CVCVCV_{SG}$→$CVCVC_{PL}$). These results constitute a conceptual replication of previous results on palatalization, where additional examples of 0→i/tʃ__ helped t→tʃi, a different operation resulting in the same product, over 0→i/t__ (the same operation resulting in a different product); Kapatsinski (2009, 2013). In both cases, ensuring that one particular structure is over-represented in forms with the meaning that the participant is asked to produce or judge leads the participant to believe that forms with that meaning should ideally fit the over-represented structure, and to judge deviations from the structure unacceptable.

This finding follows naturally from the theory that language learners pick up on what forms – and particularly forms with certain meanings – are like. When asked to produce such a form they then aim to produce the same kind of form they

[4] The Language-by-Operation interactions make the model worse according to log likelihood, $\chi^2(2)=0$, $p=1$.

have encountered serving this purpose (Bybee, 2001; Kapatsinski, 2013). When they are asked to simulate the production process in a judgment task – judging whether the plural form is the right one to produce – participants judge forms that are unlike the forms they have encountered serving this purpose unacceptable.

However, this theory is insufficient because at least some participants do learn subtraction from the short, five-minute training. This is quite clear in production data (Kapatsinski, under contract), where a substantial number of participants do subtract across the board, regardless of stimulus length. Subtraction of the final vowels of CVCV singulars cannot be preferred over addition of a consonant to such forms if one aims to produce the kinds of plurals one has experienced because all plurals the participants experienced were CVCVC plurals. With a CVCV base, such plurals can be produced by addition but not by subtraction. Subtraction cannot even be learned by associating a singular schema to a plural schema (Nesset, 2005, 2008), because this would instead result in truncation: mapping all sorts of singulars, short and long, onto a fixed prosodic template for plurals, CVCVC.

Learning subtraction appears to involve learning an *operation*, much like proposed in rule-based theories. Elsewhere, I have proposed that the operation involved is *Copy*, understood as copying an activated representation from long-term memory into the production plan (Kapatsinski, under contract), which has a strong resemblance to output-output faithfulness in Optimality Theory. Part of learning the morphology of the language is then learning what to Copy, and what not to Copy. Learning subtraction is learning not to Copy from a particular position in the source form; here, from the final position. Like a schema, Copy can be activated and inhibited by various aspects of context; here, one needs to inhibit Copying an activated final vowel of the singular form when trying to generate a plural form. In a connectionist framework, this can be modeled by having inhibitory connections from the singular-final vowel (or its phonological and positional features) and the plural meaning to Copy. Strengthening these inhibitory associations would allow one to execute subtraction across the board.

Note that learning not to Copy the final vowel involves noticing that the plural lacks the final vowel of the corresponding singular. Since order of forms was random during training, there is no guarantee that a plural will occur next to or even close to the corresponding singular. Thus, noticing the relationship between the two forms is a challenging task: one needs to keep one form in memory until the other appears, and then to retrieve the memorized form to compare with the form one perceives (e.g. Brooks et al., 1993; McNeill, 1963, 1966).[5] To the extent

---

[5] Alternatively, the learner might hypothesize the other form and keep that prediction in memory to check against the actual perceived form (Ervin, 1961; McNeill, 1963, 1966; Plunkett & Juola, 1999). Regardless, this is a memory-demanding task.

that one fails to do so, an example of $CVCVkV_{SG}$→$CVCVk_{PL}$ will not provide support for subtraction.[6]

Just as one might not be able to compare two forms of the same word, one might fail to determine the meaning of a form one hears, or fail to associate the meaning with the form. If one fails to extract 'singular' from a picture of a single creature paired with a CVCVk form, or does not think the number of creatures on the screen is important (to the choice of the form), then that example of CVCVk can provide support for producing CVCVk plurals. This would predict that CVCVk plurals benefit from examples of CVCVk paired with the *wrong* meaning. Depending on how bad the participants are at tracking particular semantic features, CVCVk plurals might be helped by additional CVCVk examples even if all such examples are paired with the wrong meaning. If inattention to meaning were severe enough, then addition of [k] would be judged as being more acceptable after exposure to the SO language than after exposure to the Baseline language. Figure 1 shows a numerical trend in that direction but it is not significant. The lack of significance is unlikely to be a power issue: Bayesian hypothesis testing using the BIC approximation to the Bayes Factor (Wagenmakers, 2007) shows that the null hypothesis is 191 times more likely than the alternative (for the PO vs. Homophones language comparison, $\Delta BIC=10.5$, $BF=\exp(10.5/2)=191$ in favor of the null). Furthermore, languages in which CVCVk was paired with the right meaning favor addition more strongly than the language in which CVCVk was always paired with the wrong meaning. These results suggest that the participants did identify and pay attention to the numerosity of creatures in pictures paired with CVCVk examples.

It is, however, the case that presenting participants with additional CVCVk examples in both the singular and the plural improves judgments of CVCVk plurals. The Homophones language affords CVCVk plurals much higher acceptability than the Baseline language. In fact, the Homophones language is not significantly different from the languages in which CVCVk appears only in plurals (Table 6). Because of the rather large sample size in the present experiment, this result provides strong evidence for the null hypothesis of no difference. Bayesian hypothesis testing using the BIC approximation to the Bayes Factor (Wagenmakers, 2007) shows that the null hypothesis is 2480 times more likely than the alternative (for the PO vs. Homophones language comparison, $\Delta BIC=15.63$, $BF=\exp(15.63/2)=2480$ in favor of the null).

This result may be somewhat surprising because the Homophones language makes CVCVk a uniquely bad cue to the plural meaning. However, this finding

[6] It may well be the case that acquisition of paradigmatic mappings such as those underlying subtraction (as well as arbitrary mappings specific to gender classes) is crucially dependent on occasionally encountering the paradigmatically related forms in close temporal proximity. At the very least, such encounters do appear to help acquisition of paradigmatic mappings in humans (Smolek & Kapatsinski, in prep), as well as help discovery of paradigmatically related forms by machines (Baroni et al., 2002; Xu & Croft, 1998).

may not indicate inattention to semantics. Instead, it may be due to schemas being production structures (Kapatsinski, 2013). In production, the form to be produced is the outcome, predicted by semantic features one wishes to express as well as the features related forms. The form to be produced is therefore naturally conditioned on the meaning and on other related forms, making the production of a form proportional to p(form|meaning). This quantity is the same in the Homophones language, the PO language and the PO+SO language, and is unaffected by what singular forms are like.

Nonetheless, these results seem to fly in the face of homophony avoidance and the empirical results offered in support of it, including the finding that languages appear to avoid mergers that collapse important semantic distinctions (Wedel et al., 2013), and that speakers hyperarticulate phonetic cues when they distinguish members of a minimal pair (Baese-Berk & Goldrick, 2009). Participants in our experiments do not appear to avoid homophony. Nonetheless, I am persuaded by these data that speakers of natural languages do avoid homophony. How then does sensitivity to homophony come about in natural language?

I would like to suggest that sensitivity to homophony is crucially dependent on experiencing feedback from the interlocutor. When one produces a form in a communicative situation, that production may either succeed or fail in communicating the intended meaning to the listener. If the listener indicates communication failure, one may learn to avoid that form in the future. In fact, precisely this kind of adjustment in response to listener feedback is documented by Buz et al. (2016), Maniwa et al. (2009), Seyfarth et al. (2016), and Schertz (2013). Goldstein & West (1999) and Goldstein et al. (2003) have also argued that pre-linguistic vocalizations are shaped into speech by the same process of reinforcement learning / instrumental conditioning. It is likely through this kind of mechanism that sensitivity to the recoverability of meaning, p(meaning|form), arises, and specific strategies to avoid misunderstanding in particular contexts by shooting for less easily confusable perceptual targets are developed (see Kapatsinski, under contract, Chapter 9, for the full story).

Future work should examine how reinforcement learning can change the process of morphological learning. If it is true that sensitivity to recoverability of meaning is not inherent to the process of inducing schemas from perceptual experience or production practice in the absence of listener feedback, then our experiments on morphology learning are missing a crucial ingredient for modeling morphology learning "in the wild" and the likely trajectories of language change (see also Kirby et al., 2008).

There is one limitation to the generality of the conclusion that learners do not automatically track p(meaning|form) in learning morphology in the absence of listener feedback. Each training trial of the present experiment (as well as in Kapatsinski, 2009, 2013) involved participants looking at a picture of a referent before hearing the corresponding wordform. Ramscar et al. (2010) have suggested that learning form-meaning associations is fundamentally directional: the association is formed from the past to the present, but not from the present to past.

If learners are presented with forms followed by pictures, and their processing resources are occupied immediately afterwards by a demanding math problem, they do not seem to learn which features of the pictures are particularly predictive of forms (picture→form associations). We should note that, in the absence of distraction, adults in Ramscar et al.'s study appeared to update probabilities in both directions, and may therefore have been expected to do so in the present experiment as well. Nonetheless, results of head camera data on children acquiring words from interactions with caregivers suggests that learning experiences conducive to word learning involve the learner looking at an object for a prolonged period of time beginning before the object is labeled – as in the present experiment – but also extending for a couple seconds after the offset of the word (Pereira et al., 2014). It may be the case that these extended fixations may be needed to provide the learner of a language with an opportunity to update both form→meaning and meaning→form associations. Future work should examine the effects of this more natural kind of presentation to determine whether sensitivity to the probability of the meaning given the form emerges under these conditions even in the absence of listener feedback on ambiguous productions.

6. Conclusion

When we try to produce a novel form of a word we know, we have access to the meaning we intend to produce and (often) to other forms of the same word. The task at hand is to express the intended meaning using the other forms of the same word as sources of raw material from which the new form can be constructed.

What is a good form to construct? The schematic speaker constructs a form that is like other forms that have the meaning she wants to express. The rule-based speaker constructs a form in the way they have seen such forms constructed, performing the same operations on the raw material of other forms. The schematic judge cares about the product, not how you get there. The rule-based judge cares about how the form is constructed, its relationship to other forms.

The present results suggest that there is truth to both approaches. Learners in this study have learned that final vowels of singulars should not be copied into the new plural form, and that plural forms should have a certain shape, CVCVk. When these two generalizations are in conflict, as they are for a CVCV singular, learners decide based on how much statistical support each generalization has.

The statistical support for the CVCVk schema is based on how many plural forms are of the CVCVk shape. It seems relatively unaffected by whether CVCVk is also encountered with other meanings. These results suggest that the goodness of a form-meaning pairing is based on the joint probability / frequency of that form-meaning pairing (Bybee, 1985, 2001), or on the conditional probability of the form given the meaning (Kapatsinski, 2013). It does not appear to be based on the extent to which the form is a good cue to the meaning, i.e. the conditional probability of the meaning given the form (or the difference in the probability of the meaning in the presence of the form vs. in its absence). The schemas that compete with operations in the present study appear to be truly product-oriented: the statistics that determine their strengths can be gathered within one cell of the morphological paradigm.

Interestingly, these product-oriented schemas emerge in languages that are better described by pure subtraction: there are many exceptions to the 'CVCVk~PL' schema, but no exceptions to the 'delete the final vowel to form the plural' rule. A rule-based grammar is better than a grammar that includes competing schemas, affording greater certainty. Yet, competing schemas are nonetheless extracted. Given how easy form-meaning mappings are to learn relative to paradigmatic form-form mappings, the learner can't help but notice that plural forms are all of the same shape.

The emergence of product-oriented schemas in these languages suggests a diachronic pathway from subtraction to truncation. In a subtraction system, the same unit is subtracted from all inputs, whatever results. In a truncation system, inputs are subtracted, augmented or left unchanged as needed to fit a particular output template, though subtraction remains the dominant operation (Inkelas, 2015). For example, in Japanese, nicknames suffixed with -tʃan are usually derived by truncating the base to two morae, but a one-mora name is lengthened instead (Poser, 1990). The present data suggest that a subtraction system in which one shape is over-represented among the outputs of subtraction carries within it the seeds of a truncation system, where that same shape is produced in a variety of ways.

References

Albright, Adam (2008). Explaining universal tendencies and language particulars in analogical change. In Jeff Good (Ed.), *Linguistic universals and language change* (pp.144-181). Oxford: Oxford University Press.

Albright, Adam, & Hayes, Bruce (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition*, *90*, 119-161.

Baese-Berk, Melissa, & Goldrick, Matthew (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, *24*, 527-554.

Baroni, Marco, Matiasek, Johannes, & Trost, Harald (2002). Unsupervised discovery of morphologically related words based on orthographic and semantic similarity. *Proceedings of the Workshop on Morphological and Phonological Learning of ACL/SIGPHON-2002*, 48–57.

Barr, Dale J., Levy, Roger, Scheepers, Christoph, & Tily, Harry J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255-278.

Bates, Douglas, Maechler, Martin, Bolker, Ben, & Walker, Steve (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1-48.

Becker, Michael, & Gouskova, Maria (2016) Source-oriented generalizations as grammar inference in Russian vowel subtraction. *Linguistic Inquiry*, *47*, 391-425.

Beckman, Jill N. (1997). Positional faithfulness, positional neutralisation and Shona vowel harmony. *Phonology*, *14*, 1-46.

Berko, Jean (1958). The child's learning of English morphology. *Word, 14*, 150–177.

Brooks, Patricia J., Braine, Martin D. S., Catalano, Lisa, Brody, Ruth E., & Sudhalter, Vicki (1993). Acquisition of gender-like noun subclasses in an artificial language: The contribution of phonological markers to learning. *Journal of Memory and Language, 32*, 79–95.

Bybee, Joan (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, Joan (1985). *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.

Ervin, Susan M. (1961). Changes with age in the verbal determinants of word-association. *American Journal of Psychology*, *74*, 361-372.

Goldstein, Michael H., King, Andrew P., & West, Meredith J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*, *100*, 8030-8035.

Goldstein, Michael H., & West, Meredith J. (1999). Consistent responses of human mothers to prelinguistic infants: the effect of prelinguistic repertoire size. *Journal of Comparative Psychology*, *113*, 52-58.

Gouskova, Maria, & Becker, Michael (2013). Nonce words show that Russian yer alternations are governed by the grammar. *Natural Language & Linguistic Theory*, *31*, 735-765.

Inkelas, Sharon (2015). *The interplay of morphology and phonology*. New York: Oxford University Press.

Kapatsinski, Vsevolod (Under contract). *Changing minds changing tools: From learning theory to language acquisition to language change*. Cambridge, MA: MIT Press.

Kapatsinski, Vsevolod (2013). Conspiring to mean: Experimental and computational evidence for a usage-based harmonic approach to morphophonology. *Language*, *89*, 110-148.

Kapatsinski, Vsevolod (2012). What statistics do learners track? Rules, constraints and schemas in (artificial) grammar learning. In Stefan Th. Gries & Dagmar Divjak (Eds.), *Frequency effects in language learning and processing* (pp.53-73). Berlin: de Gruyter.

Kapatsinski, Vsevolod (2009). *The architecture of grammar in artificial grammar learning: Formal biases in the acquisition of morphophonology and the nature of the learning task* (Doctoral dissertation, Indiana University).

Kapatsinski, Vsevolod (2005). To scheme or to rule: Evidence against the Dual-Mechanism Model and the Rule-Based Learner. *Berkeley Linguistics Society*, *31*, 193-204.

Kirby, Simon, Cornish, Hannah, & Smith, Kenny (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*, 10681-10686.

Köpcke, Klaus-Michael, & Wecker, Verena (2017). Source-and product-oriented strategies in L2 acquisition of plural marking in German. *Morphology*, *21*, 77-103.

Maniwa, Kazumi, Jongman, Allard, & Wade, Travis (2009). Acoustic characteristics of clearly spoken English fricatives. *The Journal of the Acoustical Society of America*, *125*, 3962-3973.

McNeill, David (1966). A study of word association. *Journal of Verbal Learning & Verbal Behavior*, *5,* 548-557.

McNeill, David (1963). The origin of associations within the same grammatical class. *Journal of Verbal Learning & Verbal Behavior*, *3,* 250-262.

Nesset, Tore (2008). *Abstract phonology in a concrete model*. *Cognitive Linguistics and the morphology-phonology interface*. Berlin: Mouton de Gruyter.

Nesset, Tore (2005). Opaque softening: A usage-based approach. *Poljarnyj Vestnik*, *8*, 55-68.

Pereira, Alfredo F., Smith, Linda B., & Yu, Chen (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin & Review*, *21*(1), 178-185.

Pierrehumbert, Janet B. (2006). The statistical basis of an unnatural alternation. In Louis Goldstein, Douglas H. Whalen, & Catherine T. Best (Eds.) *Laboratory phonology 8* (pp.81-107). Berlin: Mouton de Gruyter.

Plunkett, Kim, & Juola, Patrick (1999). A connectionist model of English past tense and plural morphology. *Cognitive Science, 23*, 463–490.

Poser, William J. (1990). Evidence for foot structure in Japanese. *Language, 66*, 78-105.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org/.

Ramscar, Michael, Yarlett, Daniel, Dye, Melody, Denny, Katie, & Thorpe, Kirsten (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, *34*, 909-957.

Rumelhart, David E., & McClelland, James L. (1986). On learning the past tenses of English verbs. In McClelland, James L., Rumelhart, David E., & PDP Research Group. *Parallel distributed processing* (Vol. 2). Cambridge, MA: MIT Press.

Schertz, Jessamyn (2013). Exaggeration of featural contrasts in clarifications of misheard speech in English. *Journal of Phonetics*, *41*, 249-263.

Seyfarth, Scott, Buz, Esteban, & Jaeger, T. Florian (2016). Dynamic hyperarticulation of coda voicing contrasts. *The Journal of the Acoustical Society of America*, *139*, EL31-EL37.

Smolek, Amy, & Kapatsinski, Vsevolod (In prep). The importance of contiguity for learning paradigmatic mappings. Ms. University of Oregon.

Wagenmakers, Eric-Jan (2007) A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, *14*, 779-804.

Wedel, Andy, Jackson, Scott, & Kaplan, Abby (2013). Functional load and the lexicon: Evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts in language change. *Language & Speech*, *56*, 395-417.

Xu, Jinxi, Croft, W. Bruce (1998). Corpus-based stemming using co-occurrence of word variants. *ACM Transactions on Information Systems (TOIS), 16*, 61-81.

Proceedings of the 41st annual Boston University Conference on Language Development

edited by Maria LaMendola
and Jennifer Scott

Cascadilla Press Somerville, MA 2017

Copyright information

ISSN 1080-692X
ISBN 978-1-57473-076-0 (2 volume set, paperback)
ISBN 978-1-57473-176-7 (2 volume set, library binding)

Ordering information

To order a copy of the proceedings or to place a standing order, contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, sales@cascadilla.com, www.cascadilla.com