The Inflectional Base(s) of the Russian Imperative

Andrei Antonenko

07/21/2008

1 Introduction

Reference grammars of Russian and previous literature on the imperative differ in what they take to be the base form for imperative formation. Some postulate that the imperative is formed from the basic stem of the verb, an abstract representation, from which all verb forms are derived (Jakobson 1948, Townsend 1980, Swan 1984, Es'kova 1985). Others claim that the imperative is derived from the present (non-past) tense stem, another abstract representation of the verb (Vinogradov 1972:464-5, Pulkina and Zakhava-Nekrasova 1974:251-5). According to Zaliznjak (1977:89) the basic form for the imperative is the 3pl form of the verb. Švedova (1982:620-1) argues for a large set of very specific rules deriving the imperative from the non-past stem together with the 1sg form of the verb. Others propose that in order to derive the imperative, one needs to consider two forms of the verbal paradigm. Gvozdev (1961:326-7) claims that depending on the stress and quality of the final consonant of the stem, the basic form for the imperative is either the 1sg or the 2sg form of the verb. Researchers arguing for an abstract verbal stem still need access to actual verbal forms to identify it and then there are questions of which forms are compared in deriving the basic stem and how such a stem is actually learned.

There are two main goals of this paper. The first is to test Albright's (2002) claim that inflectional paradigms may be derived from an existing paradigmatic form which can be identified as an inflectional base. I evaluate various verbal forms as potential bases for imperative formation in Russian in order to determine which form serves as the best base for the imperative. In order to do so, I apply the Minimal Generalization Learner (MGL) model (Albright and Hayes 1999) to derive the imperative form of the verb from other verbal forms, and compare the results based on number of characteristics, including the percentage of imperatives derived correctly. The Russian verbal paradigm is complex with several phonological factors, such as stress and consonantal mutations, unpredictable from any single given form of the paradigm in general. Nevertheless, I will demonstrate that based on the formation of the imperatives, each of the members of the paradigm carries statistically equal (or very close) informational load, allowing

for correct derivation of the imperatives in statistically the same number of cases; and even though the information provided by various forms differs qualitatively, it is essentially the same quantitatively. I show that learning of the imperative in Russian is guided not by attraction to any single form of the paradigm, but by the entire non-past subparadigm together. While the imperative cannot achieve 100% connections with all of non-past forms at the same time, since the non-past forms vary in their stress assignment and consonantal alternations, it can occupy the middle position within the non-past subparadigm when the connections with each of the non-past forms are statistically equal.

Under the assumption that learning is facilitated by access to more information, the second goal is to propose and test an extension of the Albright and Hayes' learnability model which allows it to take into consideration multiple bases. I suggest two ways in which multiple bases can be incorporated into the model, and test these scenarios on the Russian data to see whether they significantly increase the number of correctly derived imperatives. I then consider the predictions made by the Minimal Generalization Learner for the acquisition of the imperative forms in Russian and look at some acquisition errors.

The organization of the paper is as follows. In section 2 I briefly discuss the verbal conjugation in Russian and describe the basic rules for the formation of imperative forms. Section 3 describes the Minimal Generalization Learner, and how it is used in order to choose the base of the imperative. I also outline the problems one might expect when considering a single-base approach. Section 4 provides the results of applying the Minimal Generalization Learner to Russian imperative data. In this section I evaluate numerically and compare various choices of the basic form. Section 5 proposes an extension of Albright and Hayes' model in order to incorporate multiple bases. I describe two different options, and compare the multiple-base approach to the single-base approach. In section 6 I provide an analysis of issues identified by the Minimal Generalization Learner. Section 7 contains the discussion of the theoretical issues, as well as predictions for the acquisition of the imperatives.

2 Imperative formation: basic patterns

In the Russian verbal conjugation the present tense is inflected in two dimensions: number (singular and plural), and person (1st, 2nd, and 3rd). There are no gender differences expressed on the verb in the present tense. Because the present tense forms serve as the future

tense in perfective verbs, the "present tense" paradigm is often referred to as NON-PAST. The NON-PAST paradigm of each verb consists of six different forms. In the past tense the verb has only gender and number, but not person.

Russian verbs consist of a root, followed by an optional verbal suffix and by inflectional affixes. The root is a minimal meaningful morpheme, incorporating the main semantic features of the verb; the stem is the root together with the verbal suffix. All other forms of the verb are derived by attaching inflectional endings to stems. In general, for each verbal lemma there is always only one root, although there might be more than one stem. Traditionally Russian is said to have two stems for each verb. One stem, the *non-past stem* (NON-PAST), is used in the formation of the present tense paradigm of imperfective verbs, the simple future of perfective verbs, the present tense participles and verbal adverbs, and imperatives. The second stem, referred to as *infinitive-past stem* (INF-PAST, is used in the formation of the past tense, the past tense participles and verbal adverbs, and the infinitival form of the verb, as shown in (1).

(1)

Root:	*golos* 'to vote'					
Stems:	NON-PAST:	*golos-uj*			INF-PAST:	*golos-ova*
Present tense:	1sg	*golos-uj-u*	Past tense:	masc-sg	*golos-ova-l*	
	2sg	*golos-uj-eš*		fem-sg	*golos-ova-l-a*	
	3sg	*golos-uj-et*		neut-sg	*golos-ova-l-o*	
	1pl	*golos-uj-em*		pl	*golos-ova-l^{j}-i*	
	2pl	*golos-uj-et^{j}e*				
	3pl	*golos-uj-ut*				
Imperative:	*golos-uj*		Infinitive:	*golos-ova-t^{j}*		

The table in (2) provides the classification of Russian verbs in terms of these two stems. Vowels |i| and |e| in the NON-PAST stem are thematic vowels, and they are not relevant for imperative formation. Apart from the shape of the INF-PAST and NON-PAST stems, this table also gives the different conjugational classes with their infinitive, 2sg and imperative forms. Classes listed as productive have thousands of members, while those listed as rare are unproductive and in most cases are limited to only a few verbs (Townsend 1980).

(2) *Verbal conjugation classes (adopted from Timberlake 2004:100).*

INF-PAST	NON-PAST	*Infinitive*	*2sg*	*Imperative*	*Gloss*	*Productivity*
{CVC-i-}	{CVC-lil}	l^{j}ubí^{j}t^{j}	l^{j}úb^{j}iš	l^{j}ub^{j}í	love	productive
{CVC-e-}	{CVC-lil}	smotr^{j}ét^{j}	smótr^{j}iš	smotr^{j}í	look at	~50
{CVČ-a-}	{CVČ-lil}	molčát^{j}	molčíš	molčí	be silent	~30
{CVC-a-}	{CVC-aj-lel}	brosát^{j}	brosáješ	brosáj	toss	productive
{CVC-e-}	{CVC-ej-lel}	p^{j}jan^{j}ét^{j}	p^{j}jan^{j}éješ	p^{j}jan^{j}éj	get tipsy	productive
{CVC-ova}	{CVC-uj-lel}	tr^{j}ébovat^{j}	tr^{j}ébuješ	tr^{j}ébuj	require	productive
{CVC-nu-}	{CVC-n-lel}	brýznut^{j}	brýzn^{j}eš	brýzn^{j}i	splash	productive
{CVC-a-}	{CVC^{j}-lel}	plákat^{j}	pláčeš	pláč	cry	~60
{CVC-a-}	{CVC-lel}	sosát^{j}	sos^{j}oš	sos^{j}í	suck	rare
{CCa-}	{CC-lel}	ždát^{j}	žd^{j}óš	žd^{j}í	wait	rare
{CCa-}	{CVC-lel}	brát^{j}	b^{j}er^{j}óš	b^{j}er^{j}í	take	rare
{CVJa-}	{CVJ-lel}	davát^{j}	dajóš	dáj	give	rare
{CVJa-}	{CVJ-lel}	kl^{j}evát^{j}	kl^{j}ujóš	kl^{j}új	peck	rare
{CV-}	{CVJ-lel}	žít^{j}	živ^{j}óš	živ^{j}í	live	rare
{CV-}	{CVJ-lel}	krýt^{j}	króješ	krój	cover	rare
{CV-}	{CJ-lel}	p^{j}ít^{j}	p^{j}jóš	p^{j}éj	drink	rare
{CV-}	{CVN-lel}	d^{j}ét^{j}	d^{j}én^{j}eš	d^{j}én^{j}	set	rare
{CV-}	{CN-lel}	žát^{j}	žm^{j}óš	žm^{j}í	squeeze	rare
{CVRV-}	{CVR-lel}	kolót^{j}	kól^{j}eš	kol^{j}í	prick	rare
{CVR(V)-}	{CR-lel}	um^{j}er^{j}ét^{j}	umr^{j}óš	umr^{j}í	die	rare
{CVC-}	{CVC-lel}	n^{j}est^{j}í	n^{j}es^{j}óš	n^{j}es^{j}í	carry	rare

The imperative form in Russian is possible in the 2sg and the 2pl form. The 2pl form of the imperative is unambiguously predictable based on the 2sg imperative by the addition of the suffix *-t^{j}e* to the 2sg imperative form: the verb *xod^{j}i* 'walk-2sg.imp' yields the 2pl imperative *xod^{j}i-t^{j}e*. Since this process is entirely predictable I focus only on the formation of the 2sg form of the imperative.[1]

[1] There is also a 1pl imperative form, but morphologically this form is always the same as the 1pl non-past form (with *-t^{j}e* optionally added at the end of it), and a synthetic 3sg form of the imperative which is formed by the verb *pust^{j}* 'let' and the 3sg non-past form of the verb: *pust^{j} čitajet* 'read-3sg-imp'.

In order to form the imperative, a speaker needs to choose between two allomorphs: -∅ and *-i*, and to identify a stem or base form. The choice of -∅ versus *-i* depends on the sonority sequencing of consonant clusters and prosody: *-i* is preferred if the stem ends in a consonantal cluster of rising sonority, e.g *kr*^{j}*íkn*^{j}*i* 'shout', and if the stem is not stressed, e.g. *p*^{j}*iší* 'write'. Information about the segmental structure of the stem is available in any verbal form, but information about the stress pattern is only available in the 1sg (or, to some extent, in the infinitive).

The stress of the 1sg non-past form on the verb is partly correlated with the stress of the infinitive: if in the infinitive the stress falls on the verbal suffix, it will fall either on the verbal suffix or on the ending in the 1sg non-past form of the verb: *čitát*^{j}*, čitáju* 'read-inf, 1sg' or *p*^{j}*isát*^{j}*, p*^{j}*išú* 'write-inf, 1sg'. If the stress in the infinitive falls on the verbal root, the root will also be stressed in the 1sg non-past form of the verb (and in all other forms): *plákat*^{j}*, pláču, pláčeš* 'cry-inf, 1sg, 2sg'. Timberlake (2004) distinguishes several different stress patterns in the NON-PAST tense. These patterns are exemplified in (3) below: (3a) represents a fixed stress pattern, where the stress consistently falls on the root or on the conjugational suffix/thematic vowel/inflectional ending in all verbal forms; (3b) exemplifies a mobile stress pattern, where stress shifts between the 1sg inflectional affix and the syllable preceding the thematic vowel/inflectional ending in other forms.

(3) a. Fixed stress pattern: Stress on the root

pláč-u, pláč-eš 'cry-1sg, 2sg'

slávl^{j}*-u, sláv*^{j}*-iš* 'glorify-1sg, 2sg'

Fixed stress pattern: Stress on the conjugation suffix (if syllabic) / thematic vowel / inflectional ending

čit-áj-u, čit-áj-eš 'read-1sg, 2sg'

n^{j}*es-ú, n*^{j}*es*^{j}*-óš* 'carry-1sg, 2sg'

b. Mobile stress pattern: Stress shifts between the 1sg inflectional affix and the syllable preceding the thematic vowel / inflectional ending in other forms

p^{j}*iš-ú, p*^{j}*íš-eš* 'write-1sg, 2sg'

The existence of the mobile stress pattern (3b) is crucial: it shows that the inflectional ending of the 1sg non-past form (or the verbal suffix in the infinitive) can carry stress, while all other verbal forms are stressed on the preceding syllable (root or stem). Therefore, the position of the stress in the 1sg non-past form and the infinitive in general cannot be predicted from the other forms in the paradigm. In order to choose the imperative allomorph, it is necessary to take the 1sg non-past form (or the infinitive) of the verb into consideration, as shown in (4).

(4) *The 1sg non-past has stress on the inflectional ending:*

Infinitive	***1sg non-past***	***2sg non-past***	***Imperative***	***Gloss***
t^{j}an-út^{j}	t^{j}an-ú	t^{j}án^{j}eš	t^{j}an^{j}í (*t^{j}án^{j})	'pull'
p^{j}is-át^{j}	p^{j}iš-ú	p^{j}íšeš	p^{j}iší (*p^{j}íš)	'write'

If the 1sg form of the verb is stressed on the stem, then the imperative allomorph is -∅ if this choice does not result in a violation of the Sonority Sequencing Principle (SSP), as in (5a).[2] Otherwise, the *-i* suffix is chosen, as in (5b).

(5)

	Infinitive	***1sg non-past***	***Imperative***	***Gloss***
a.	*The 1sg non-past stressed on stem, no SSP violations occur:*			
	plák-at^{j}	pláč-u	pláč (*pláči)	'cry'
	ob^{j}íd^{j}et^{j}	ob^{j}íž-u	ob^{j}íd^{j} (*ob^{j}íd^{j}i)	'offend'
	čitá-t^{j}	čitáj-u	čitáj	'read'
b.	*The 1sg non-past stressed on the stem, stem ends in illegal cluster:*			
	kr^{j}íkn-ut^{j}	kr^{j}íkn-u	kr^{j}íkn^{j}i (*kr^{j}íkn^{j})	'shout'
	zam^{j}édl-it^{j}	zam^{j}édl^{j}-u	zam^{j}édl^{j}i (*zam^{j}édl^{j})	'slow down'

[2] In general, Sonority Sequencing Principle violations are tolerated in Russian: *igr* 'game-gen.pl', *m^{j}etr* 'meter'. However the formation of imperatives shows that even though low ranked, this principle/constraint is nevertheless active.

While the choice of the imperative allomorph can be made solely based on the 1sg form of the verb, this form is not sufficient to generate imperatives, as the following problem arises. Consider the examples in (6).

(6)			***1sg***	***Imperative***	***2sg***	***Gloss***
	a.	i.	l^{j}ubl^{j}ú	l^{j}ub^{j}í	l^{j}úb^{j}iš	'love'
		ii.	skobl^{j}ú	skobl^{j}í	skobl^{j}íš	'scrape'
	b.	i.	l^{j}ečú	l^{j}et^{j}í	l^{j}et^{j}íš	'fly'
		ii.	l^{j}ečú	l^{j}ečí	l^{j}éčiš	'heal'

In the examples (6a-i) and (6b-i) the final stem consonants in the 1sg form and the imperative differ: *bl*^{j} vs. *b*^{j} in (6a-i) and *č* vs. *t*^{j} in (6b-i). Such consonantal alternations (mutations) are not predictable from the 1sg form. Final stem consonants in the 1sg of the verbs in (6a-ii) and (6b-ii) are preserved in the imperative, and do not undergo alternations. Example (6b) presents the extreme case when the 1sg forms of the verbs 'to heal' and 'to fly' are homophonous, but the imperatives are different. For verbs in (6) the necessary information about consonantal alternations is available in the 2sg form: the final stem consonant in the imperative is the same as the final stem consonant in the 2sg form. Thus knowing only the 1sg form is not sufficient to derive the imperative. Detailed discussion of consonantal mutations in Russian is provided in section 3. Given that information about stress and consonantal mutations relevant to imperative formation often cannot be obtained from any single form of the verb, it appears that several bases need to be considered.

An alternative approach to verbal conjugation was developed by Jakobson (1948), who proposes that all forms of the verbal paradigm can be derived from a single verbal base. This basic stem of the verb is also marked for the stress pattern associated with it: whether it is a mobile pattern or not. The basic stem of the verb serves as a base for all other verbal forms, including the infinitive. This presupposes that speakers can somehow establish a basic stem for every verb. The main problem with an analysis of imperatives based on the notion of a basic stem of the verb is the problem of learnability. No account of how learners arrive at the basic stem of the verb exists, and clearly the only input learners get consists of free-standing forms of

verbs. Even though derivation of the imperative from the basic stem is straightforward, identifying the basic stem itself is not, and this task is not necessarily easier than deriving the imperative from the other basic forms of the verb directly, without resorting to the basic stem.

3 Choice of the base for Russian imperatives

In this section I show how learners can derive the imperative form of the verb without going through the intermediate step of identifying the basic stem of the verb. I will test the Minimal Generalization Learner, originally developed by Albright and Hayes (1999) based on the model outlined by Pinker and Prince (1988), and later described in Albright 2005. This is a rule-based model of rule discovery, which analyzes dependencies between one form in the paradigm ("base") and another ("output"), and generates a set of morpho-phonological rules which can be used to derive the output from the base.

The basic description of the minimal generalization model of rule induction is the following. The model starts by generating word specific rules for each pair of input-output, which are subsequently generalized by comparing rules producing the same change, and constructing other, more general rules, referring not to particular lexical items, but to environments in which certain changes occur. For example, a hypothetical situation in trying to derive the imperative from the 1sg non-past form of the verb is as follows. Assume the inputs and outputs as in (7).

(7) a. xočú → xotʲí
b. lʲečú → lʲetʲí
c. močú → močí

The learner starts with word specific rules similar to those in (8).

(8) a. čú → tʲí / xo ___ based on (7a)
b. čú → tʲí / lʲe ___ based on (7b)
c. čú → čí / mo ___ based on (7c)

These rules would be enough if no generalization were needed. However, learning grammar means learning to derive the necessary generalizations. Therefore, ultimately the learner will compare the possible rules and come up with more general versions of them, such as the rules in (9).

(9) a. čú → tʲí / V ___ comparing (8a) and (8b)

b. čú → čí / V ___ generalizing (8c)

Each of the rules has a numerical characteristic, *reliability*, associated with it. Reliability is calculated by dividing the number of the forms included in the rule's structural change (=hits) by the number of forms included in the rule's structural description (=scope). In the previous example, for instance, both rules (9a) and (9b) apply for all three inputs from (7), therefore the scope of these two rules is equal to 3. The rule in (9a) produces the correct output in 2 cases, therefore its reliability equals 2/3. The rule in (9b), however, is correct for only one input and has a reliability of 1/3.

Reliability ratios are adjusted using lower confidence limit statistics to yield *confidence values* (see Albright 2005, Mikheev 1997 for details). Such adjustment is necessary for a relative comparison of the rules. For example, a rule which is not 100% correct but makes correct predictions for 990 inputs out of 1000 will have a higher confidence value than the rule making a correct prediction in 2 out of 2 values, even though the unadjusted reliability of the former rule (990/1000=0.99) is lower than the unadjusted reliability of the latter (2/2 = 1.00). Thus, rules which cover fewer forms have lower confidence values when compared to rules which are more general. When new forms are derived, the rule with the highest reliability is applied, and the output obtained by using this rule becomes an actual output.

I compiled a list of the 531 most frequent regular Russian verbs, together with all six present tense (non-past) forms, the infinitive, and the imperative from the online frequency dictionary by Sharoff, which was created on the basis of a corpus of modern Russian and contains a selection of texts from both fiction and non-fiction written during the last quarter of the twentieth century. Verbs used have a frequency of more than 50 instances per million. Stress and palatalization of consonants were marked in each of the forms.

Several verbs were excluded from the data used for testing (10).

(10) a. *jést*^{j} 'to eat', and all prefixed forms of it

b. *dát*^{j} 'to give', and all prefixed forms of it

c. *jéxat*^{j} 'to ride, to go', and all prefixed forms of it

d. verbs with the prefix *vý-*

Verbs (10a) and (10b) are irregular and have suppletive 1sg non-past forms, as well as an irregularly formed imperative. The verb in (10c) exhibits irregular consonantal mutation: in the non-past forms of this verb the consonant [x] alternates with the consonant [d], while in the imperative it alternates with [ž:]. Such irregular change does not occur in any other verbs, and therefore the paradigms of these verbs cannot be generalized to any other verbs. The problem with verbs prefixed with *vý-* is different in nature. This prefix, being the only stressed verbal prefix in Russian, affects the formation of imperatives. If this prefix is attached to a verb which takes *-i* as a surface realization of the imperative suffix, the prefixed form will also show *-i* in the imperative form, as in (11a). Since *vý-* always carries a stress, the 1sg non-past form of the verb is not stressed on its inflectional ending, and according to the regular rules described above, the imperative should not be formed with the *-i* suffix but with -∅. The situation is even more interesting if the non-prefixed imperative form of the verb has the ∅-allomorph. In such cases, the *vý*-prefixed form of the verb allows for two alternatives in the formation of the imperative: one with the overt *-i* and one with -∅, as shown in (11b).

(11)

	Infinitive	***Imperative***	***Gloss***
a.	kur^{j}ít^{j}	kur^{j}í	'smoke'
	výkur^{j}it^{j}	výkur^{j}i	'smoke out'
b.	brós^{j}it^{j}	brós^{j}	'toss'
	výbros^{j}it^{j}	výbros^{j}/výbros^{j}i	'throw out'

The reason to exclude such verbs from testing is related to the local nature of the Minimal Generalization Learner: the model only derives rules based on a local environment. Since this prefix is not in a local environment with the imperative ending, the model will not be able to take its presence into account. Recent work by Albright and Hayes (2006) mentions a new model of the Learner able to deal with non-local environments, but the code of a non-local learner is not yet available.

The Java-based version of the Minimal Generalization Learner, available from Bruce Hayes' website (http://www.linguistics.ucla.edu/people/hayes/learning/), was executed on the data with the task of learning how to derive the imperative form based on the non-past tense forms and the infinitive. In fact, only four forms out of seven (six non-past forms and the infinitive) are sufficiently different. The 2sg, 3sg, 1pl, and 2pl always have the same base and differ only in the inflectional affix. Therefore, I considered the derivation of the imperative from the following four forms: 1sg, 2sg, 3pl, and the infinitive.

There were two reasons to include the infinitive form: 1) Acquisition errors often include imperative forms derived from the INF-PAST form as in *p^{j}isáj* (children's form) from *p^{j}isát^{j}* (Inf) instead of expected *p^{j}išť* 'write' or *r^{j}isováj* (children's form) from the infinitive *r^{j}isovát^{j}* instead of *r^{j}isúj* 'draw'; also, Bar-Shalom and Snyder (1999) mention that children acquiring Russian often use the infinitive form of the verb in place of the imperative, and 2) to numerically evaluate whether the contribution of the infinitive to one of the non-past forms produces significant improvement in the number of imperatives derived correctly, i.e. can the infinitive serve as a second base needed to derive the imperative, or do both bases need to be finite? As far as I am aware, this problem has not been addressed in the previous literature. I answer this question in my discussion of the model's extension to several bases in section 5.

The data for the Learner consisted of the four verb forms (Inf, 1sg, 2sg, 3pl), represented in (12) below:

(12)

	Inf	***1sg***	***2sg***	***3pl***	***Imp***	***gloss***
a.	móč	mogú	móžeš	mógut	mog^{j}í	'can'
b.	skazát^{j}	skažú	skážeš	skážut	skaží	'tell'
c.	čitát^{j}	čitáju	čitáješ	čitájut	čitáj	'read'

d.	smotr^{j}ét^{j}	smotr^{j}ú	smótr^{j}iš	smótr^{j}at	smotr^{j}í	‘watch’
e.	žal^{j}ét^{j}	žal^{j}éju	žal^{j}éješ	žal^{j}éjut	žal^{j}éj	‘have pity’
f.	r^{j}isovát^{j}	r^{j}isúju	r^{j}isúješ	r^{j}isújut	r^{j}isúj	‘draw’
g.	l^{j}ub^{j}ít^{j}	l^{j}ubl^{j}ú	l^{j}úb^{j}iš	l^{j}úb^{j}at	l^{j}ub^{j}í	‘love’
h.	otv^{j}ét^{j}it^{j}	otv^{j}éču	otv^{j}ét^{j}iš	otv^{j}ét^{j}at	otv^{j}ét^{j}	‘answer’

There is a great deal of variability in consonant alternations and stress among different forms of the verb. For instance, the verb in (12a) has a stem-final consonant which alternates between four different possibilities: [č, g, ž, g^{j}]. Similar, although less drastic, consonant alternations can be observed in the verbs in (12b), (12g), and (12h). Further, the difference between the verbs in (12b) and (12c) illustrates the problems which arise while trying to derive the imperative from the infinitive. Both of these verbs have a C-final root; however, the verb in (12c) has the suffix *-aj-* in the NON-PAST stem, and the verb in (12b) does not. Such a difference can only be captured by considering one of the non-past forms of a verb, since the suffix *-aj-* only surfaces there. A similar difference distinguishes verbs in (12d) and (12e), with the verb in (12e) having the suffix *-ej-* in the non-past forms of the verb. Example (12f) presents a suffix alternation between *-ova-* in the INF-PAST forms of the verb and *-uj-* in the NON-PAST forms. Also, the stress patterns are different for the verbs. The verbs in (12a), (12b), (12d), and (12g) have mobile stress which falls on the inflectional ending in 1sg, Inf, and Imp forms, while the verbs in other examples have fixed stress on either the stem or the inflectional ending.

Verb forms with their stress and consonant alternations were entered in a spreadsheet, and the Minimal Generalization Learner was run four times with the following inputs and output.

(13) a. 1sg → Imp
b. 2sg → Imp
c. 3pl → Imp
d. Inf → Imp

Since the code has the ability to generate rules based on the phonetic features of the segments, the necessary data about the basic features of Russian segments were entered in the table and provided to the learner. The following features were provided: ±consonantal, sonority (rated on the scale from 0 to 5, with 5 corresponding to vowels, and 0 to stops), ±voice, ±LAB, ±COR, ±DORS, and ±palatalized (for consonants), height, and backness (for vowels).

A further assumption of Albright and Hayes' model is the following: learners of language are familiar with phonotactic constraints in the language and are therefore able to correct outputs violating phonotactic constraints. I compiled a list of phonotactically illegal sequences in Russian and the code of the Minimal Generalization Learner had access to it. The list included the following items: coda clusters violating sonority sequencing (where an obstruent is followed by a sonorant), such as *bl*, *bn*, *br*, *bj*, etc.; sequences of a non-palatalized consonant followed by a high front unrounded vowel *i*, such as *bi*, *ni*, *ji*, etc.; sequences of palatalized labial or velar obstruents followed by a high back rounded vowel *u*, such as $b^{j}u$, $k^{j}u$, $f^{j}u$, etc.; and palatalized velar obstruents in word-final position: $k^{j}\#$, $g^{j}\#$ and $x^{j}\#$. Even though this list is far from being exhaustive, these sequences constitute the phonotactically illegal sequences to be avoided in imperatives.[3]

There are two predictions: 1) Given the structure of the Russian imperative and its relation to the NON-PAST paradigm, the Minimal Generalization Learner should have a much lower success rate for the infinitive (PAST-INF) than for finite (NON-PAST) forms; 2) Given that the necessary information about stress and CC-sequences is found in the 1sg form, this form should be more successful in generating the correct imperative than other NON-PAST forms.

There were four executions of the code in an attempt to derive an imperative from the 1sg, 2sg, 3pl, and the infinitive forms of verbs. The algorithm generated a set of rules which can account for imperative formation from the above mentioned forms, and each rule was assigned a

---

[3] Sequences violating the sonority sequencing principle are often allowed in Russian, especially in the nominal paradigm, e.g *rúbl*^{j} 'ruble', t^{j}*ígr* 'tiger' and the ban on such sequences in imperatives is possibly a result of the interaction between faithfulness and markedness constraints: nominal forms such as *rubl*^{j} 'ruble' surface because correcting a violation of the sonority sequencing principle would require either epenthesis of a vowel, or deletion of a stem consonant, while the sonority sequencing problem in imperatives can be easily solved by choosing an appropriate imperative allomorph. In the OT framework this difference can be captured if the faithfulness constraints against epenthesis or deletion are ranked higher than the markedness constraint against violation of sonority sequencing principle.

confidence value showing how likely this rule is to be used in imperative derivation. Wug-tests were also conducted in order to evaluate the correctness of the generated rules, and following the learning/rule-generation state, imperative forms were derived for all verbs used as the input to the Minimal Generalization learner. Since some inputs satisfy the environment requirements of multiple generated rules, a substantial number of the inputs yielded multiple possible forms for the imperative. In this particular test, the maximal number of generated possible outputs from one input was equal to three (see (14a,b)).

(14) The cases of more than one outputs from one input:

a. 3pl → Imperative; input *prʲedlóžat* 'offer-3pl', actual output *prʲedložít*

(i) *prʲedlóž* (ii) *prʲedlóži* (iii) *prʲedložít*

b. 2sg → Imperative; input *uxódʲiš* 'leave-2sg', actual output *uxodʲí*

(i) *uxódʲ* (ii) *uxódʲi* (iii) *uxodʲí*

c. 1sg → Imperative; input *proisxožú* 'happen-1sg', actual output *proisxodʲí*

(i) *proisxodʲí* (ii) *proisxožú*

d. Infinitive → Imperative; input *nablʲudatʲ* 'observe-inf', actual output *nablʲudaj*

(i) *nablʲudʲí* (ii) *nablʲudaj*

The examples in (14) provide several representative situations; the detailed discussion of the individual problematic cases is given in Section 8. The cases in (14a), (14b), and (14c) are representative of two common learning mistakes that arise if the base is chosen to be one of the NON-PAST forms of the verb. The choice of the imperative suffix depends on the stress in the 1sg form of the verb: if it is stressed on the stem, the imperative is realized by a palatalized final consonant of the stem, otherwise, if it is stressed on the affix, the imperative allomorph is *-í*. Such information about stress is not readily available in the 2sg and 3pl forms, and therefore the large number of 2sg and 3pl inputs yield at least two outputs, one with *-i* and one without, such as cases in (14a,b-i) and (14a,b-ii). Further, the Learner failed to predict whether stress in the imperative falls on the ending like in (14a,b-iii) or on the stem, as in (14a,b-ii), as both cases are possible in Russian. Note that the learner did not generate any outputs where stress occurs on a

syllable other than the imperative ending or in its original position. For instance, in the case of (14a), there were no generated forms with stress on *e*, such as *pr^{j}édlož* or *pr^{j}édloži*. This result is expected, as none of the input-output pairs exhibits such a stress alternation.

A different problem can be seen in the derivation of the imperative from the 1sg form. Stress is not an issue here: all the generated outputs in example (14c) are stressed on the inflectional affix: when the 1sg form has affix stress in the input the surface form of the imperative marker (which is based on stress) is chosen correctly. What is a problem here are the stem alternations, which occur very often in the 1sg form. The table in (15) identifies the consonants in Russian which undergo so called "substitutive palatalization", or "mutation", (as opposed to "bare palatalization", which does not involve any major segmental changes, but only adds the [+palatalized] feature to the consonant, changing *s* to *s^{j}*, *t* to *t^{j}*, etc.).

(15) Consonant mutations in Russian:

Labials	Dentals	Velars
p → pl^{j}	t → č	k → č
b → bl^{j}	d → ž	g → ž
f → fl^{j}	s → š	x → š
v → vl^{j}	z → ž	
m → ml^{j}		

Examples of verbs exhibiting mutations are given in (16).

(16)

	Infinitive	*1sg*	*2sg*	*3pl*	*Imperative*	*Gloss*	*Pattern*
a.	Labial mutations:						
i.	l^{j}u**b^{j}**ít^{j}	l^{j}u**bl^{j}**ú	l^{j}ú**b^{j}**iš	l^{j}ú**b^{j}**at	l^{j}u**bí**	'love'	1
ii.	t^{j}er**p^{j}**ét^{j}	t^{j}er**pl^{j}**ú	t^{j}ér**p^{j}**iš	t^{j}ér**p^{j}**at	t^{j}er**p^{j}**í	'tolerate'	1
b.	Dental mutations:						
i.	pr^{j}á**t**at^{j}	pr^{j}á**č**u	pr^{j}á**č**eš	pr^{j}á**č**ut	pr^{j}á**č**	'hide'	2
ii.	xo**d^{j}**ít^{j}	xo**ž**ú	xó**d^{j}**iš	xó**d^{j}**at	xo**d^{j}**í	'walk'	1
iii.	p^{j}i**s**át^{j}	p^{j}i**š**ú	p^{j}íše**š**	p^{j}íše**š**ut	p^{j}iší	'write'	2
iv.	vo**z^{j}**ít^{j}	vo**ž**ú	vó**z^{j}**iš	vó**z^{j}**at	vo**z^{j}**í	'drive'	1

c. Velar mutations:

i.	plá**k**at^{j}	pláču	pláčeš	pláčut	pláč	'cry'	2
ii.	napr^{j}a**g**át^{j}	napr^{j}a**g**ú	napr^{j}a**ž**óš	napr^{j}a**g**út	napr^{j}a**g^{j}**í	'strain'	3

d. Exceptional patterns:

i.	xo**t^{j}**ét^{j}	xočú	xóčeš	xo**t^{j}**át	xo**t^{j}**í	'want'	4
ii.	b^{j}e**ž**át^{j}	b^{j}e**g**ú	b^{j}e**ž**íš	b^{j}e**g**út	b^{j}e**g^{j}**í	'run'	5

Examples in (16) demonstrate various patterns which occur in Russian with respect to consonantal mutation. The labial mutation, as shown in (16a), only occurs in the 1sg form of the verb and changes labial consonants to labial consonants followed by palatalized *l*. I denote this pattern by 1 in the column "Pattern". Knowing that the 1sg form of the verb ends in a labial consonant followed by *l^{j}u* does not by itself guarantee that this sequence is obtained by mutation. Consider for instance the verb with the infinitive form *skobl^{j}ít^{j}* 'scrape'. All finite forms and the imperative of this verb will have a *bl^{j}* sequence: *skobl^{j}ú*, *skobl^{j}íš*, *skobl^{j}át*, *skobl^{j}í* 'scrape-1sg, 2sg, 3pl, imp'. This pattern is also possible with dental mutations. The relevant examples are given in (16b-ii, iv). The consonants *d* and *z* surface as mutated *ž* in only the 1sg non-past form, and do not mutate in any other verbal forms.

However, this is not the only pattern possible with dentals. Pattern 2 in examples (16b-i, iii) involves mutations in all non-past tense forms, as well as in the imperative. There the infinitive is the only form under consideration which does not exhibit mutation.

Velars have two distinct patterns. The verb in (16c-i) exemplifies pattern 2. Pattern 3, where mutated consonants appear only in 2sg, 3sg, 1pl, and 2pl forms is shown in (16c-ii). This pattern is impossible with labials or dentals.

Verbs in (16d) show two exceptional mutation patterns. Pattern 4 exemplified in (16d-i) involves mutations in 1sg, 2sg, and 3sg and occurs only with the verb 'want' and its prefixed counterparts. In pattern 5 mutated consonants occur in 2sg, 3sg, 1pl, 2pl forms, and the infinitive. This pattern is only possible with the verb 'run' and its prefixed counterparts.

Getting back to the multiple outputs generated by the Minimal Generalization Learner, we now can see what the problem is with example (14c). The final stem consonant undergoes

mutation $d^j \rightarrow \check{z}$ in the 1sg form (Pattern 1), and therefore the Learner generates two possibilities: one preserving the mutated consonant $\check{z}$, and one having the bare palatalized non-mutated consonant d^j.

A common problem identified by the Learner in the derivation of the imperative from the infinitive form is demonstrated in (14d). As can be seen from the table of verbal conjugation classes in (2) and discussion of examples (10b-e), the NON-PAST stem can include either the *j*-ending suffixes *-aj-*, *-ej-* or vowel suffixes *-a-* and *-e-*. The infinitive gives no indication of which suffixes occur in the NON-PAST stem, as suffixes *-aj-* and *-a-* both correspond to infinitives ending in *-at*^{j}, and suffixes *-ej-* and *-e-* correspond to *-et*^{j} ending infinitives; only the finite forms of the verb provide information about them. Therefore in trying to derive the imperative from the infinitive, the Learner does not have valid information regarding the presence of such suffixes, and thus the two possible outputs cover both alternatives.

There are other, less common problems, which I will not consider here.

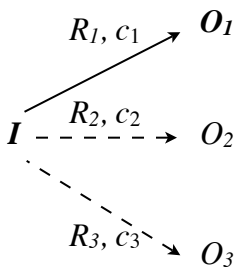
4 Quantificational analysis

A quantificational analysis should be able to show whether any of the verbal forms considered above fares better than any other form for generating imperatives, based on a number of parameters. The five parameters considered in this analysis are the following.

1) The percentage of imperatives derived correctly. Recall that the Minimal Generalization Learner in general derives several outputs from one input by applying different rules depending on the given environment. Each rule has its confidence value, which is equal to the ratio of the number of correct outputs derived by this rule to the total number of cases where the rule can be applied, statistically adjusted using the method described by Mikheev (1997). The larger the confidence value is, the more dependable the rule is. Here I assume that the surface form of the imperative generated by the Learner for each input is an output which is derived by the rule which has maximal confidence. For instance, if the input *I* yields outputs O_1, O_2, and O_3 by rules R_1, R_2, and R_3 with respective confidences c_1, c_2, and c_3, such that $c_1 > c_2 > c_3$, the output chosen by the Learner is O_1, since it is generated by the rule with the maximal confidence. Schematically this situation is shown in (17), where dashed arrows represent the

derivations using rules with confidence values lower than maximal confidence, and the chosen output O_1 is given in bold.

(17)



The percentage of imperatives derived correctly was computed using wug-tests conducted after the learning stage, and is equal to the ratio of correct outputs generated by the learner to the total number of verbs under consideration. This parameter is the main characteristic demonstrating how good of a predictor an input form is.

2) The average number of outputs. As mentioned earlier and shown in the schematic representation of the learning model in (17), the Learner generates several outputs for each input. This characteristic provides the average number of outputs generated from each input. The smaller this number is, the less variability is observed in the outputs, and therefore the computational load required to derive the output is less since there is no competition between the outputs. Thus, the input form which has a lower number of outputs must be superior to the input form with a higher average number of outputs.

3) The average confidence of correct outputs. This characteristic is equal to the average confidence value of the rules which derive the outputs (imperative forms) occurring in the language. If the Learner failed to correctly generate an output at all, and none of the generated outputs are occurring, the confidence was considered to be equal to 0. The meaning behind this characteristic is accuracy of the input as a predictor for imperatives. Numerically, this value ranges from 0 to 1, and the closer it is to 1, the higher are the confidence values of the rules deriving correct outputs. If the value is close to 0, it indicates that the real language forms are either not generated by the Learner or are generated by rules having very low confidence values.

4) The average confidence value of winning rules. The output which the model chooses as the best one is the output generated by the rule which has maximal confidence. This

characteristic is equal to the average of confidence values of the winning rules. Notice that the large value of this characteristics does not necessarily mean that the corresponding form is a better predictor for the imperative, as one can imagine that the generated outputs might be wrong, but still have very high confidence values. What it means is that the particular verbal form is in a sense very "self-confident".

5) The average difference between confidence values of the winning rule and its closest competitor. For each input form the difference between the confidence values of the rule with the highest confidence and of the closest competing rule was calculated. In case only one output was generated, the closest competing rule was assumed to have a confidence value equal to 0. To obtain this parameter, the differences were further averaged over the set of all verbs. The larger this characteristic is, the larger the margin by which the output of the model wins against its competitors, and the less serious the competition for a given input is.

In order to evaluate which of the forms of the verb serves as a better predictor for the imperative, I provide the distributions of these parameters below.

Figure 1 shows the percentage of imperative forms generated/predicted correctly out of total number of 531 verbs used in the test. Even though the absolute differences between the input forms are not large, the 2sg, the 3pl forms (each at about 93%) and the 1sg form (95%) fare better in comparison to the infinitive (87%). ANOVA was applied to the data, and showed the effect of the choice of the input form on the percentage of the imperatives derived correctly ($F_{3,2120} = 6.865$; $p < 0.001$). Further, Bonferroni post-hoc tests showed that the percentages of correct predictions from finite inputs (1sg, 2sg, 3pl) do not differ significantly from each other, while the number of correct

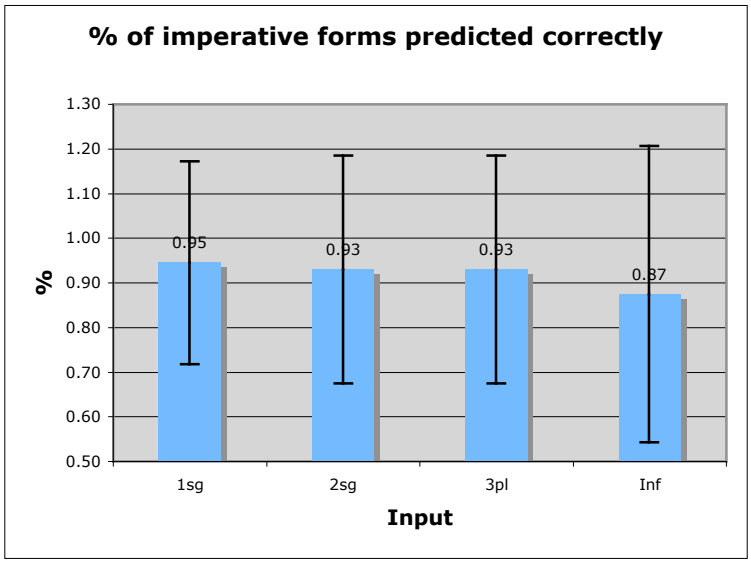


Fig. 1. The percentage of imperative forms predicted correctly.

outputs from the infinitive is significantly different from the numbers obtained by using finite forms as inputs. Therefore, based on this parameter, no differences exist among non-past forms of the verb: all of them are equally good in generating the imperative. The infinitive however shows significantly lower performance in deriving the imperative. The performance of the infinitive in predicting the imperative is not surprising if we recall that the infinitive is derived from the PAST-INF stem, while the imperative, as well as non-past forms of the verb, are all derived from the NON-PAST stem of the verb.

Figure 2 shows the average number of possible outputs generated from one input. As can be seen, the average number of outputs ranges from 1.51 when the 3sg form is used as a base for the imperative up to 1.80 when the infinitive form is used. ANOVA results show that the effect of the input is significant ($F_{3,2120}$ = 18.064; $p < 0.001$). Post-hoc tests reveal that the differences in this parameter between infinitive and finite forms are significant, while the differences between the various finite forms are not.

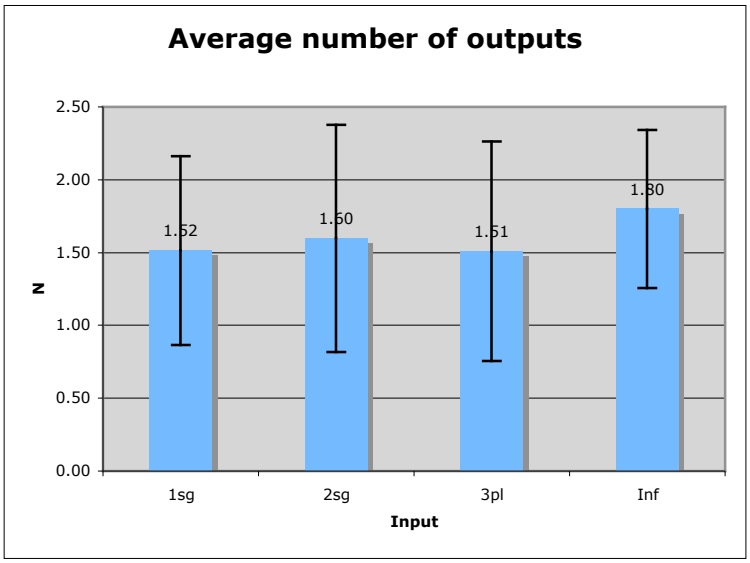


Fig. 2. The average number of outputs.

Among the finite forms, a closer look at the standard deviation shows that the 2sg and 3pl forms have a larger standard deviation, σ=0.78, and σ=0.75 respectively, when compared to that of the 1sg base, σ=0.65. Therefore, based on this statistic, 2sg and 3pl forms fare worse than the 1sg form because of too much uncertainty in the choice of imperative for some forms. In order to understand the source of this uncertainty, let us look at the distribution of the number of outputs depending on the form used as base. This data is provided in Figures 3a and 3b.

Figures 3a and 3b give the same data in two different representations. Figure 3a shows the percentage of one, two, and three outputs for different input forms. Figure 3b groups the same data by the number of outputs. As can be seen from these graphs, the infinitive yields two

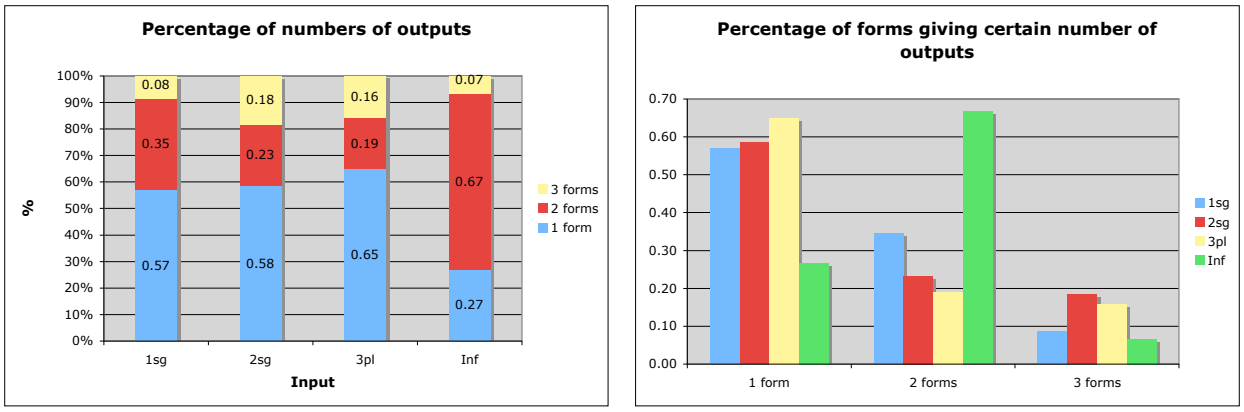


Fig. 3a, b. Distribution of the number of outputs.

outputs in 67% of the cases compared to 35%, 23%, and 19% for 1sg, 2sg, and 3pl, respectively. The finite forms in the majority of cases generate only one output (57%, 58%, and 65% for 1sg, 2sg, and 3pl, respectively), while the infinitive yields one output in 27% of the cases. Even though the 3pl form generates one output in a maximal number of cases, it also yields three outputs in 16% of the cases, which is comparable to 18% for the 2sg, and is much higher than 8% and 7% for the 1sg and the infinitive, respectively. Why is this so? Recall that neither the 2sg nor the 3pl forms provide the necessary information about the prosody of the verbal base, and therefore cannot justify the choice of the surface form of the imperative marker or the position of the stress in the imperative (see examples (12a) and (12b)), yielding a high number of three-outputs. The infinitive and the 1sg forms provide stress information, but lack either morphological information about the presence of a *j*-ending suffix (in case of the infinitive) or phonological information about the consonantal changes (in the 1sg form). Thus, these forms rarely give rise to more than 2 outputs.

The distributions of the next two parameters are given in Figures 4 and 5.

With respect to the average confidence values for occurring output forms, the ANOVA results show that the effect of the input form is significant ($F_{3,2120} = 16.376$; $p < 0.001$). Further, post-hoc tests reveal that the infinitive, with the average confidence of the correct output equal to 0.78, fares significantly worse than the other forms (which have average confidence values of correct forms ranging from 0.84 to 0.87), while there is no significant difference among the finite forms. These results indicate that the occurring forms receive lower confidence values when derived from the infinitive as opposed to the finite forms.

The next parameter, the average confidence of the winning rule, is presented in Figure 5. ANOVA finds a significant effect of the input form on this parameter as well ($F_{3,2120} = 14.661$; $p < 0.001$). From post-hoc tests it is clear that the infinitive is the least "self-confident" form. There is a significant difference between the infinitive and any other finite form. Further, the 3pl form is significantly different from the 1sg; however, its difference with the 2sg form is not significant according to the post-hoc tests. Also, the difference between the 1sg and the 2sg forms is not significant.

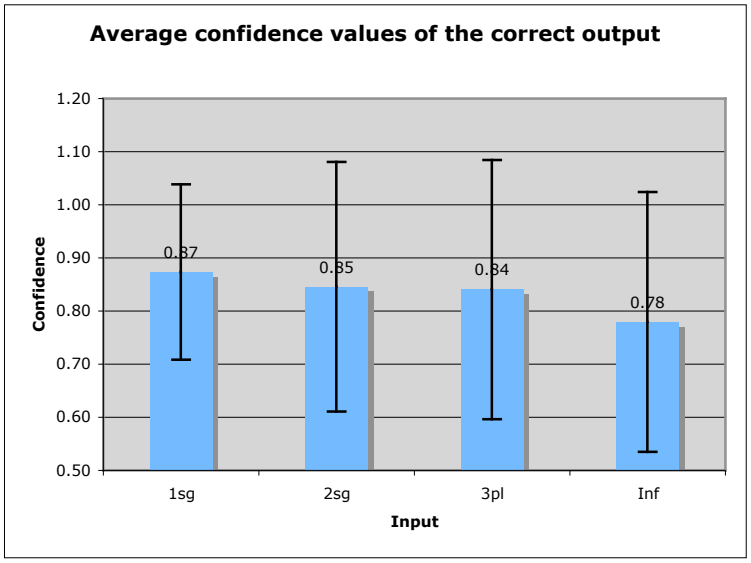


Fig. 4. The average confidence values of the occurring forms.

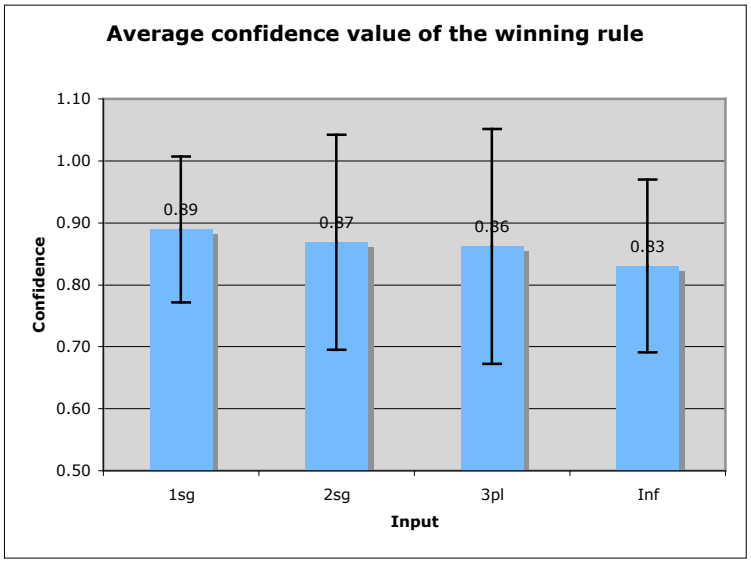


Fig. 5. The average confidence value of the winning rule.

The last parameter, the average difference between confidence values of the winning rule and its closest competing rule, is given in Figure 6. ANOVA shows a significant effect of the input on this parameter ($F_{3,2120} = 16.712$; $p < 0.001$). Further, pairwise comparisons using post-hoc tests reveal significant difference between the infinitive and all other forms, while there are no significant differences between the finite forms. The significantly lower value of this parameter for the infinitive indicates that competing output forms obtained from it are significantly closer to each other than outputs from any other input form under consideration.

The final Figure 7 in this section shows the percentage of imperative forms derived correctly from 0 (never), 1, 2, 3, or from any of the 4 input forms under consideration (1sg, 2sg, 3pl, Inf). From this graph it is clear that the overwhelming majority (79%) of imperatives are derived correctly regardless of which verbal form is chosen as input. At the same time, 1% of all imperatives cannot be derived correctly no matter which form of the verb is chosen as base. 13% of the imperatives are derived correctly from 3 out of 4 forms, and only 3% of the imperatives are derived correctly from 1 and from 2 input forms.

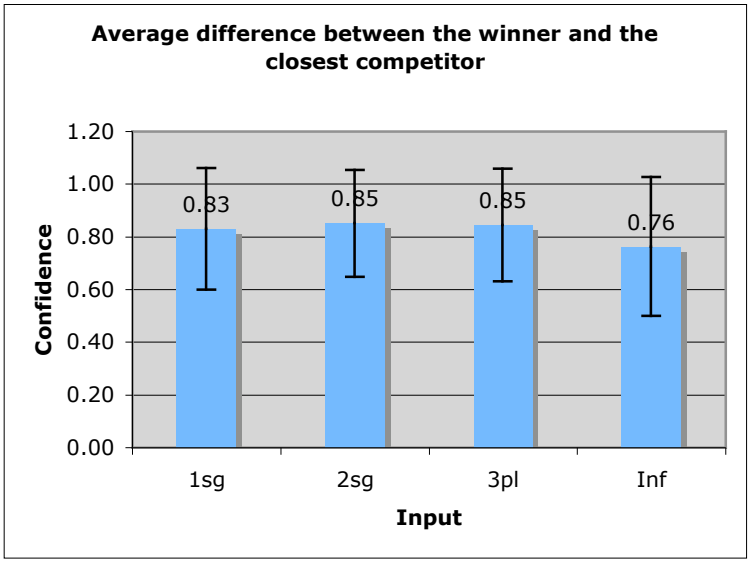


Fig. 6. The average difference between the confidence values of the winning rule and its closest competing rule.

So, were the statistical data computed based on the results of Minimal Generalization Learner surprising? Both yes and no. As predicted, infinitives do not fare as well in predicting imperatives in comparison to finite non-past forms of the verbs, as there are several major pieces of information not present in the INF-PAST stem. Surprising however are two facts. First is the fact that even though the infinitive lags behind the finite forms, it is still fairly powerful in predicting the correct imperative form, performing successfully in 87% of the cases. Second,

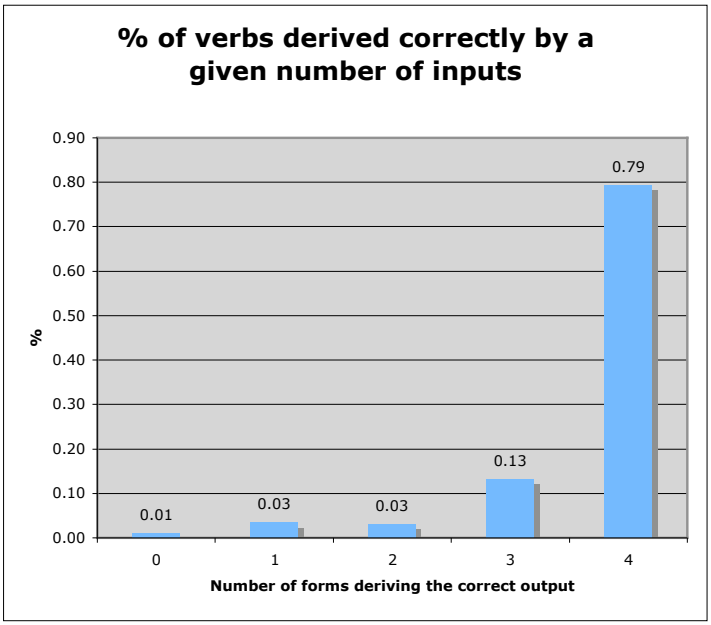


Fig. 7. Percentage of verbs derived correctly from a given number of inputs.

surprising is the fact that no statistically significant differences were found between 1sg, 2sg, and 3pl forms except with respect to the average confidence of the winning rule parameter, and that one form, in spite of some traditional claims that two are needed, fares relatively well in predicting imperatives correctly for 93-95% of the cases. The high percentage of correct imperative derivations from any finite form is a frequency effect. I show that the problematic verbs span entire conjugational classes which have only a few members and/or are infrequent. Detailed discussion of imperatives that cannot be successfully derived from a single basic form is provided in section 6. But are multiple bases significantly better than single bases in predicting the imperatives?

5 Towards a multiple bases approach

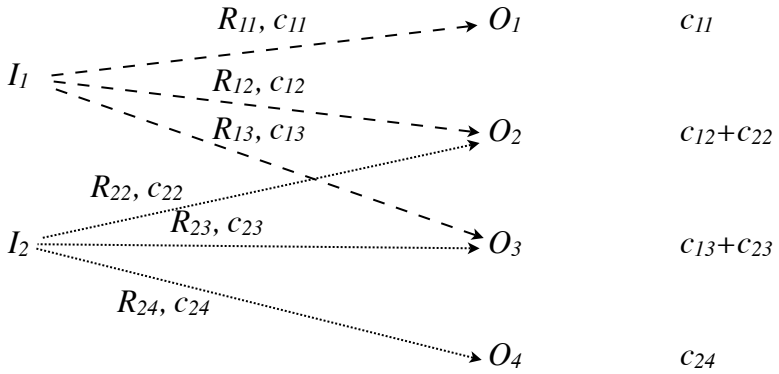
The analysis developed in the previous sections used one existing form of the verb (either one of the non-past forms or the infinitive) to derive the imperative. As I showed, the finite forms of the verb fare better than the infinitive for this task: they provide the correct output in 93% to 95% of the cases, while the infinitive yields the correct output for only 87% of all the verbs under consideration (Fig. 1). The following question therefore arises: what happens if the learner has access to several bases while trying to generate the imperative? One might expect a significant increase in rate of successful imperative derivations.

However, this question is not easy to answer. What one needs is a precise understanding of what it means to have access to two (or even more) base forms. In this section I propose two potential strategies which can be used to incorporate more than one basic form, and explore if these strategies give the Learner any benefits in the derivation of the imperative.

5.1 Scenario 1: Simultaneous derivations from multiple bases

For the first strategy, I will assume that the learner has already acquired the non-past tense forms of the verbs along with the infinitive. What that means is that the learner is able to apply the rules for the formation of the imperative not just from one base form, but from several base forms belonging to the same verbal paradigm. The confidences for obtaining the output form from different bases are then added together, and the sum of these confidences provides the final score for the given output form. Such a strategy is demonstrated schematically in (18).

(18)



Considering that the learner has access to two inputs I_1 and I_2 from the same verbal paradigm, there are several forms which can be generated. In the case from (18) the input I_1 generated three outputs O_1, O_2, and O_3 using rules R_{11}, R_{12}, and R_{13} with confidences c_{11}, c_{12}, and c_{13}, respectively. Similarly, the input I_2 generated three outputs O_2, O_3, and O_4 using rules R_{22}, R_{23}, and R_{24} with confidences c_{22}, c_{23}, and c_{24} respectively. In this hypothetical example only outputs O_2 and O_3 are generated from both I_1 and I_2; the other outputs are generated from only one of the input forms. The confidence values of the rules yielding all of the outputs are added together, so the value corresponding to output O_1 is c_{11}, O_2 is $c_{12}+c_{22}$, etc. The winning output is the form for which the sum of the confidence values is maximal.

In order to check whether this strategy provides an increase in the percentage of correctly derived outputs, the sums of the confidence values were calculated for each pair of input bases (1sg+2sg, 1sg+3pl, etc.) and for all four input bases (1sg+2sg+3pl+Inf), and the output with the largest sum of confidence values was considered to be the winner for given combination of bases. After that, the percentage of imperatives predicted correctly was calculated. The data is given in Figure 8, where the percentage of predicted forms from single bases is also given for comparison (same as in Fig. 1).

As previously, ANOVA was conducted in order to investigate the effect of the input form on the percentage of correct imperative derivations. The effect was found to be significant ($F_{10,5830} = 10.836$; $p < 0.001$). Further, post-hoc tests were conducted in order to investigate whether considering sums of the confidence values provides a significant improvement over the one-base model. As we saw before, among derivations involving one base, the differences in correct predictions of imperatives were significant between any finite form and the infinitive,

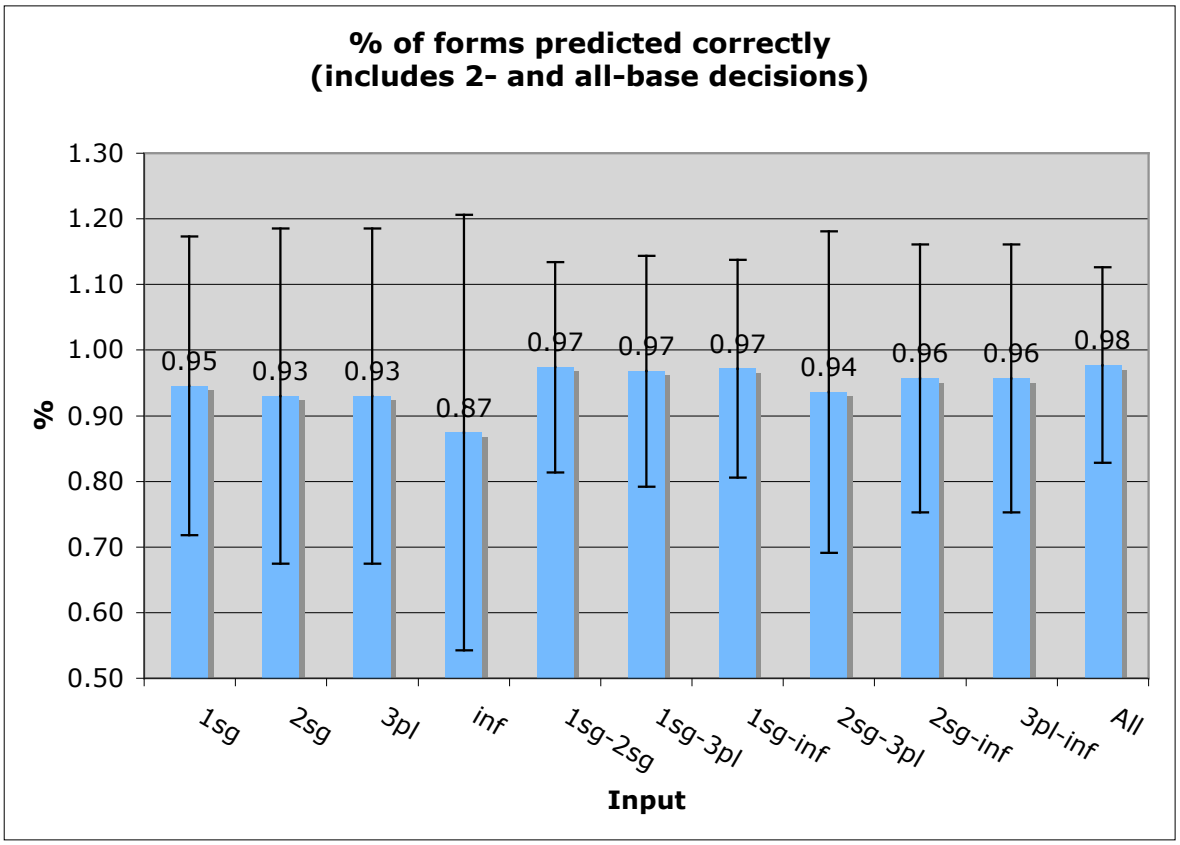


Fig. 8. Percentage of imperative forms predicted correctly (including two- and all-base decisions).

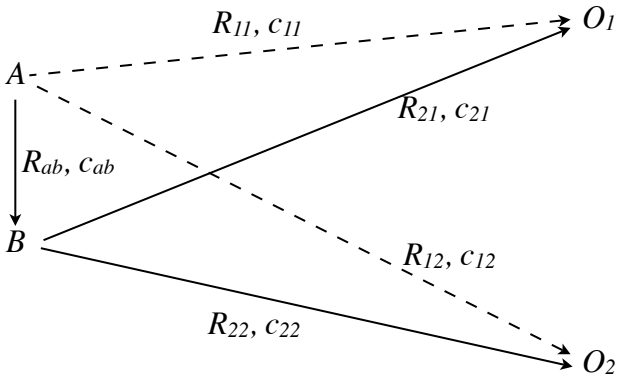
while the differences between finite forms were not significant. Starting from the 1sg form, significant improvement was achieved by combining it with the 2sg form or with the infinitive, while combining the 1sg and the 3pl forms did not significantly improve results. Further, in comparison to the 2sg form alone, combining it with the 1sg form or with the infinitive produced significantly better results, while combining the 2sg with the 3pl form did not lead to significant effects. Also, combining the 3pl with the 1sg or with the infinitive improved the imperative formation in comparison to considering the imperative derivation from the 3pl form alone, but combining it with the 2sg form did not. As expected, the infinitive benefitted from being combined with any of the finite forms. Lastly, the combination of all four input forms fares significantly better in deriving the imperative than any form considered alone, and better than the combination of 2sg and 3pl forms. However, it does not provide significantly better results than any of the remaining combinations (1sg+2sg, 1sg+3pl, 1sg+Inf, 2sg+Inf, 3pl+Inf).

The conclusion of this exercise is the following. Being able to use the rules deriving imperatives from two basic forms simultaneously and to combine the results afterwards produces significant improvement in the accuracy of imperative formation. Also, there is no need to resort to more than two forms in order to generate the imperative form. Finally, all pairs except for the combination of the 2sg and the 3pl forms fare equally well in the task of generating the imperative, and they produce results which do not differ from results obtained by combining all four forms.

5.2 Scenario 2: Indirect derivations

The second scenario I will consider in this section is the scenario when the learner tries to use intra-paradigmatic relations, establishing a connection between forms within the paradigm, to derive the imperative. The intuition behind this scenario is the following. Assume that input form *A* when used to derive the imperative by itself leads to either ambiguous results, deriving imperatives O_1 and O_2 with approximately equal confidence values, or derives a wrong imperative. Does knowing relations between the input form A and input form B, i.e. knowing the rules deriving B from A, help the learner to resolve ambiguity and to make a correct choice of imperative? The schematic illustration of this model is given in (19).

(19)



$\max(c_{11}, c_{ab}\times c_{21})$

$\max(c_{12}, c_{ab}\times c_{22})$

Here, by using the form *A* alone, there are two possibilities for the imperative: O_1 and O_2, which are obtained with confidence values c_{11} and c_{12}, respectively. Assume now that the learner knows the rules of generating another form *B* in this verbal paradigm, which is derived from *A* (correctly or not) with the confidence value c_{ab}. Now when it comes to the formation of the imperative, the learner has a choice: either to derive it directly from *A*, by following the dashed

arrows in (19), or to derive it through the intermediate form *B* by following the solid arrows in (19), using paths $A \rightarrow B \rightarrow O_1$ or $A \rightarrow B \rightarrow O_2$. Deriving the form O_1 from *A* directly gives it a confidence of c_{11}, and deriving the form O_2 from *A* directly, gives it a confidence of c_{12}. If the forms are derived through the intermediate form *B*, the confidence values of the derived imperative forms would be the products of confidence values of deriving *B* from *A* by the confidence values of deriving the imperative form from *B*. Thus if the imperative is derived through the intermediate form *B*, the confidence value of O_1 would be $c_{ab} \times c_{21}$, and the confidence value of O_2 would be $c_{ab} \times c_{22}$. Now for each of the output forms O_1 and O_2 the learner would choose an optimal way (in terms of the confidence value) of derivation (either directly or through the intermediate form), and therefore the final confidence values of the output forms will be $\max(c_{11}, c_{ab} \times c_{21})$ for O_1, and $\max(c_{12}, c_{ab} \times c_{22})$ for O_2.

If the scenario described above affects the accuracy of imperative formation, it would be possible to claim that knowing and learning intra-paradigmatic relations and the rules of deriving one verbal from another can help to establish stronger relations with other verbal forms. It would also mean that other relations within the verbal paradigm are more robust than those between a given form and the imperative, and that they can be learned more easily than the relation between a member of the paradigm and the imperative.

To test this hypothesis, I applied the Minimal Generalization Learner to all pairs of forms within the verbal paradigm, with a task to learn rules for deriving the 1sg, the 2sg, the 3sg, and the infinitive from any other form within the paradigm (total of 12 executions of the Minimal Generalization Learner code). The results of these derivations were subsequently used to generate imperatives, also by using the Minimal Generalization Learner. All twelve scenarios are given in (20).

(20)

a. 1sg → 2sg → Imp	b. 2sg → 1sg → Imp
1sg → 3pl → Imp	2sg → 3pl → Imp
1sg → Inf → Imp	2sg → Inf → Imp
c. 3pl → 1sg → Imp	d. Inf → 1sg → Imp
3pl → 2sg → Imp	Inf → 2sg → Imp
3pl → Inf → Imp	Inf → 3pl → Imp

The results of this analysis did not confirm the claim. No statistically significant improvement in accuracy of imperative formation was found when considering the indirect derivations of imperatives through intermediate forms: none of the scenarios in (20) provided a significant improvement in the number of imperatives derived correctly. Improvement was rarely seen for some individual verbs, but for some others the incorrect imperative formed indirectly had a higher confidence value than the correct imperative formed without intermediate forms: in these cases, the output form was incorrect. On average, as many verbs suffered from the indirect approach as benefitted.

This exercise demonstrated that learning the rules for deriving verbal forms used as inputs from other verbal forms (1sg, 2sg, 3pl, infinitive) is not necessarily easier than learning the rules for the derivation of imperatives. Furthermore, the amount of information which can be learned by the Minimal Generalization Learner by considering direct imperative derivation cannot be improved if the Minimal Generalization Learner also learns other intra-paradigmatic relations.

In this section I looked at the possible ways of extending the Albright model to allow it to incorporate multiple bases. I proposed two scenarios of how multiple bases can be involved in deriving the imperative form. The first scenario assumes that multiple bases are already learned, and that imperatives can be derived from all of them simultaneously, and then the results are compared to yield a joint winner. I demonstrated that under this scenario, the accuracy of imperative formation can be significantly improved. Under the other scenario, where rules which allow derivations of all members of the paradigm from all other members of the paradigms are learned by the Minimal Generalization learner (as the basic forms are not assumed to be all known), no significant improvement in accuracy was found.

6 Analysis

The quantificational analysis reveals two things: 1) any non-past form of the verb (not just the 1sg) is highly successful in deriving the imperative form, producing the correct imperative form for 93-95% of verbs, 2) the infinitive base, as expected, is not as successful as non-past forms in deriving imperatives; however, the infinitive does have a fairly high success

rate. The following questions thus arise: What makes the infinitive as successful as it is given that it is built on a INF-PAST stem? In which specific cases do certain bases fail and why?

6.1 Infinitive as base

We saw above that the choice of the infinitive as a base for the imperatives is the worst one. Only in 87% of the cases is the imperative derived correctly. This is significantly less than the percentage of imperatives derived correctly from other finite verb forms. Nevertheless, 87% is still unexpectedly high considering that the infinitive and the imperative are based on different stems. I examine reasons for the failure of the infinitive as base, which also provide an explanation for its unexpected relative success.

The number of different problems with postulating the infinitive as a base is large.

1) -at^{j} verbs. The infinitives which end in *-at^{j}* are ambiguous: the NON-PAST stem can have *-aj-* or not (see the table in (2)). Therefore, the learner may make a mistake in classifying the verb based on its infinitive since it does not contain any information about whether *-aj-* is going to surface in the NON-PAST. This is the case for the verbs in (21a-c), which belong to the *-a-* class. The verb in (21d) is a member of the *-aj-* class, and its imperative was generated correctly.

(21)

	Infinitive	***Imperative***	***Generated form***	***Gloss***	***Class***
a.	kr^{j}ičát^{j}	kr^{j}ičí	*kr^{j}ičáj	‘shout’	*-a-*
b.	spát^{j}	sp^{j}í	*spáj	‘sleep’	*-a-*
c.	plákat^{j}	pláč	*plákaj	‘cry’	*-a-*
d.	čitát^{j}	čitáj	čitáj	‘read’	*-aj-*

The class of *-aj-* verbs (21d) is productive with thousands of members. The number of verbs with the suffix *-a-* is much less as there are slightly over 100 such verbs. Therefore, the occurrence of incorrect imperatives for *-a-* type verbs (such as those in (21a-c)) is a frequency effect. However since the members of the *-a-* class of verbs are still very common and occur frequently, the incorrect imperative is derived with a confidence only slightly higher than the confidence of the correct imperative (0.59 vs. 0.57 for *kr^{j}ičat^{j}*). The verb in (21c) presents a

more complicated problem. Apart from the possibility of choosing *-aj* vs. not choosing *-aj* for the imperative, there is also mutation in the imperative.

Note that we might expect to observe similar problems with verbs, whose infinitive in *-et*^{j} does not show the NON-PAST distinction between *-ej-* and *-e-* suffixes, as in (22a) and (22b).

(22)	***Infinitive***	***Imperative***	***Gloss***	***Class***
a.	p^{j}jan^{j}ét^{j}	p^{j}jan^{j}éj	'get drunk'	*-ej-*
b.	smotr^{j}ét^{j}	smotr^{j}í	'see'	*-e-*

However, there are no incorrectly derived imperatives in these cases, even though Townsend (1980) lists the *-ej-* class of verbs as productive compared to the *-e-* class of verbs. Why is this the case? Even though the class of *-ej-* verbs is productive, the frequencies of its individual members are extremely low: only a few made the list used in this particular experiment (recall that the list included all verbs with frequencies of above 50 instances per million). Therefore, the correct imperative form of the verb in (22b) was generated, and there were no results such as **smotr^{j}éj*.

2) -ova- verbs. The next problem arises with respect to verbs whose infinitives contain the suffix *-ova-*, given in (23a) and (23b). The reflexive *-s^{j}a* suffix is ignored for present purposes.

(23)	***Infinitive***	***Imperative***	***Generated form***	***Gloss***
a.	volnovát^{j}(s^{j}a)	volnúj(s^{j}a)	*volnováj(s^{j}a)	'worry'
b.	int^{j}er^{j}esovát^{j}(s^{j}a)	int^{j}er^{j}esúj(s^{j}a)	*int^{j}er^{j}esováj(s^{j}a)	'be interested'

The suffix *-ova-* is replaced with the suffix *-uj-* in forming the non-past tense forms and the imperative. Even though the learner managed to learn to replace this suffix if it is not stressed (for instance, the Learner correctly derived the imperative *tr^{j}ébuj* from the infinitive *tr^{j}ébovat^{j}* 'request' with confidence of 0.93, and the incorrect form *tr^{j}ébovaj* only had a confidence of 0.53), it was problematic for the Learner to derive the imperative when this suffix is stressed,

even though the difference in confidence values was minimal in this case, e.g. the correct form for (23a) *volnúj(sʲa)* was derived with the confidence value of 0.72, while the confidence of the ungrammatical form *volnovájsʲa)* was 0.76. Problems with this suffix arise from the following fact. Consider the verbs in (24).

(24)		***Infinitive***	***Imperative***	***Gloss***
	a.	nalʲivátʲ	nalʲiváj	'pour'
	b.	nazyvátʲ	nazyváj	'call'

These verbs have *-va-* as a part of the stem, and their imperative ends in *-j*. Based on these verbs, the Learner generates the rule which replaces *-tʲ* at the end of the infinitive with *-j* after *-vá-* to form the imperative. Since the frequency of verbs such as the ones in (24) is high, this rule, even though it produces incorrect results for *-ová-* verbs, still has a higher confidence value than a rule that replaces *-ová-* with *-új-*. Therefore, the incorrect imperatives are derived.

3) Consonantal differences. The imperative forms of verbs in (25) all have an extra consonant which is not present in the infinitive.

(25)		***Infinitive***	***Imperative***	***Generated form***	***Gloss***
	a.	podnʲátʲ	podnʲimʲí	*podnʲáj	'raise'
	b.	načátʲ	načnʲí	*načáj	'begin'
	c.	žítʲ	živʲí	*ží	'live'

For examples (25b) and (25c) the learner could not predict the occurrence of this extra consonant, and the correct forms received confidence values close to zero. In the case of (25a) the frequency of verbs with *-m-* present in NON-PAST and absent in INF-PAST allowed the Learner to generate the correct form, but it lost to the ungrammatical form (confidence values of 0.64 for the ungrammatical form vs. 0.53 for the grammatical).

4) Vowel differences. Similar to the problem of consonantal differences, this next problem concerns vowel occurrence in the imperative.

(26)		***Infinitive***	***Imperative***	***Generated form***	***Gloss***
	a.	brát^{j}	b^{j}er^{j}í	*bráj	'take'
	b.	um^{j}er^{j}ét^{j}	umr^{j}í	*um^{j}er^{j}í	'die'
	c.	pozvát^{j}	pozov^{j}í	*pozváj	'call'

From examples (26a)-(26c) one can see that the infinitive form of the verb either has a stem vowel which is not present in the imperative or lacks a stem vowel which must be present in the imperative.

5) Remaining verbs. The remaining verbs which are problematic for the learner are verbs similar to those in (27).

(27)		***Infinitive***	***Imperative***	***Generated form***	***Gloss***
	a.	ub^{j}ít^{j}	ub^{j}éj	*ub^{j}í	'kill'
	b.	p^{j}ít^{j}	p^{j}éj	*p^{j}í	'drink'

Even though this class of verbs is small, it faces competition from verbs which are phonologically similar to the verbs in question. For instance, the ungrammatical imperative form *p^{j}í* derived from the infinitive *p^{j}ít^{j}* 'to drink' (27b) is similar to the verb *nastu**p^{j}ít^{j}*** 'to step on' whose imperative is *nastu**p^{j}í*** (common parts are bolded).

Detailed examination of the problematic cases above also reveals why the infinitive had a fairly high success rate. All problems described above, except the *-ova-* infinitives, affect nonproductive and rare classes (according to the table in (2)). The class of *-ova-* infinitives is productive, but 1) not all verbs belonging to it are affected, and 2) the frequencies of the verbs in this class are on average lower than the frequencies of the verbs in other productive classes. Therefore, verbs belonging to these problematic classes account for only about 13% of a total of 531 verbs considered in the experiment, which explains why the remaining 87% of imperatives was derived correctly.

6.2 1sg as base

The 1sg form of the verb is the best predictor when it comes to deriving imperatives. Based on this form alone, the accuracy of imperative formation is about 95%. However, the remaining 5% of the verbs present a problem for imperative derivation. Where do these problems come from? As we know, the choice of the imperative allomorph (*-i* or $-^{j}$/ palatalization) can be predicted based on the stress of the 1sg form of the verb, and therefore this choice should not be a major problem in deriving the imperative. The results of the learning model confirm this prediction. In virtually no cases was the imperative allomorph chosen incorrectly (except in two cases, discussed below). However, the problem with the derivation of the imperative from the 1sg form of the verb is due to consonant mutation. Representative cases of incorrect imperatives are given in (28).

(28)

	1sg	***Imperative***	***Generated form***	***Gloss***
a.	xo**č**ú	xo**t**^{j}í	*xo**č**í	'want'
b.	s^{j}i**ž**ú	s^{j}i**d**^{j}í	*s^{j}i**ž**í	'sit'
c.	pu**šč**ú	pu**st**^{j}í	*pu**šč**í	'let'
d.	l^{j}u**bl**^{j}ú	l^{j}u**b**^{j}í	*l^{j}u**bl**^{j}í	'love'
e.	kor**ml**^{j}ú	kor**m**^{j}í	*kor**ml**^{j}í	'feed'
f.	poznakó**ml**^{j}u	poznakó**m**^{j}	*poznakó**ml**^{j}i	'introduce'

The cases in (28) involve mutation of the final consonant of the stem, where the 1sg form has a mutated version of the consonant, while the imperative form has a non-mutated consonant (bolded in the examples). These cases correspond to Pattern 1 shown in (16) and discussed above. Further, it is interesting to note that all of the problematic cases involving the mutation of dentals (as opposed to labials) have affixal stress in the 1sg form. No problems are encountered for the verbs in (29), although they also show a mutation in the 1sg form.

(29)

	1sg	*Imperative*	*Generated form*	*Gloss*
a.	v^{j}íž**u**	v^{j}í**d**^{j}	v^{j}í**d**^{j} (not *v^{j}íž)	'see'
b.	otv^{j}éč**u**	otv^{j}é**t**^{j}	otv^{j}é**t**^{j} (not *otv^{j}éč)	'answer'

Since the learner makes predictions based on the input consisting of forms appearing in the language to explain this finding, we need to look at the possible imperative forms. In fact, there are many verbs whose stems end in *č* or *ž*, and these consonants, not being a result of a mutation, are preserved in all verbal forms, as in the verbs in (30):

(30)

	1sg	*Imperative*	*Gloss*
a.	l^{j}e**č**ú	l^{j}e**č**í	'heal'
b.	l^{j}e**ž**ú	l^{j}e**ž**í	'lie'

The presence of a [*č*], for example, in the 1sg form of the verb may correlate with either [*č*] or [*t*] elsewhere in the paradigm. The vast majority of verbs like those in (30), where there is no mutation, chose *-i* as the imperative allomorph. The incorrectly derived imperative forms in (28a-c) are constructed based on analogy with verbs such as those in (30). Furthermore, verbs with imperatives ending in *-ž* or *-č* are less frequent in the language. No similar analogy is possible for verbs in (29) and therefore their imperatives are derived correctly. In fact, the confidence values of the problematic examples (the derived incorrect form) are very close to those of the correct form. For instance, the confidence value of the correct imperative from (28a) *xot^{j}í* is 0.77, which is only 0.02 lower than the confidence value of the incorrectly derived imperative **xočí*, whose confidence value is 0.79. But the confidence value of a form like **v^{j}íž* from (29a) is 0.21, much lower than the confidence value of the correct imperative *v^{j}íd^{j}*, which is 0.72. Thus, these frequency effects explain what initially appears to be a stress effect.

Because of frequency effects, not all mutations are equal. The mutation *s* → *š* does not present problems because the frequency of the verbs with this mutation is higher than the frequency of the verbs without it. The imperative form corresponding to the 1sg form *nošú* 'I

carry' is predicted to be *nosʲí*, and not **nošú* (with the corresponding confidence values of 0.87 and 0.59 respectively).

Similarly, not all labials fare equally. (28d) and (28e) present problems with the mutation of the segments *m* and *b*, where the mutation is incorrectly preserved in the derived imperative. The difference in confidence values is not large: 0.61 vs 0.57 for incorrect and correct imperatives, respectively. The problem does not occur with other labial consonants undergoing mutation in the 1sg form, such as *p* and *v*. For instance, for the 1sg form *kuplʲú*, 'I will buy', the correct imperative *kupʲí* was derived with the confidence 0.85, while the confidence of the incorrect imperative form **kuplʲí*, which preserves the mutation, was only 0.61.

Example (28f) is interesting in the sense that it is one of the very few examples where the imperative allomorph is chosen incorrectly. However, there is a clear reason for this. As I showed above, the learner has a problem with mutated stems ending in *ml* and attempts to preserve the mutation in the imperative. The Learner tried to generate the imperative in *ml*, but because this is a phonotactically bad coda for Russian imperatives, an extra *-i* was added to solve the problem.

The major problem with choosing the 1sg form as a base for imperatives comes from the existence of mutation pattern 1, where the mutated consonant surfaces only in the 1sg form of the verb. Because this mutation pattern is impossible for velar stems, their imperatives are always derived correctly from the 1sg form.

6.3 2sg and 3pl as base

Choosing the 2sg or the 3pl form as a base for the imperative presents different problems. The incorrectly derived imperatives for the most part overlap if the base is taken to be 2sg or 3pl, and therefore these two potential bases are considered together.

The problem with the mutation in labials is now nonexistent, since that mutation only occurs in the 1sg form of the verb. The only potentially problematic cases related to mutation are the ones where the mutation occurs in the 2sg form of the verb, but does not occur in the imperative (mutation pattern 3).

(31)		*2sg*	*3pl*	*Imperative*	*Generated form*	*Gloss*
	a.	mó**ž**eš	mó**g**ut	mo**g**^{j}í	mo**g**^{j}í (not *mo**ž**í)	‘to be able to’
	b.	l^{j}á**ž**eš	l^{j}á**g**ut	l^{j}á**g**	*l^{j}á**ž**	‘lie down’
		tol**č**óš	tol**k**út	tol**k**^{j}í	*tol**č**í	‘bray’
		napr^{j}a**ž**óš	napr^{j}a**g**út	napr^{j}a**g**^{j}í	*napr^{j}a**ž**í	‘strain’

The contrast between (31a) and (31b) can be explained in terms of frequency. No problems arise in (31a) and the imperative is produced correctly. Since the verb in (31a) is very frequent, its imperative is easily learned by the Minimal Generalization Learner: the correct imperative has a confidence value of greater than 0.6, while its closest competitor is below 0.5.

On the contrary, for the verbs in (31b) the learner assumed that the imperative preserves the mutated consonant. The verbs in (31b) exhibit mutation pattern 3, where 2sg, 3sg, 1pl, and 2pl have mutations. The difference in confidence values here was also large: for instance, the incorrect form *napr^{j}aží* was generated with the confidence of 0.96, while the correct imperative lagged seriously behind with the confidence value close to zero. However, the number of such examples is very limited, and therefore it is not a major problem for deriving the imperative from the 2sg non-past form. The problem does not arise if the base of the imperative is assumed to be the 3pl non-past form of the verb as it does not contain the mutated consonant (pattern 3 does not exhibit mutation in the 3pl form).

Verbs of exceptional mutation patterns 4 and 5 are given in (32).

(32)		*2sg*	*3pl*	*Imperative*	*Generated form*	*Gloss*
	a.	xó**č**eš	xo**t**^{j}át	xo**t**^{j}í	xo**t**^{j}í (not *xo**č**í)	‘want’
	b.	b^{j}e**ž**íš	b^{j}e**g**út	b^{j}e**g**^{j}í	*b^{j}e**ž**í	‘run’

Because the verb in (32a) is very frequent, the Learner had no problems learning it and deriving a correct imperative. The verb in (32b) has a lower frequency and presented the same problems for the Learner as verbs in (31b).

The major problem in deriving imperatives from the 2sg or 3pl non-past forms is the lack of information about prosody. Thus we would expect to find cases where the imperative allomorph is chosen incorrectly. Such predictions are confirmed. All verbs exhibiting the mobile stress pattern (when the 1sg non-past form is stressed on inflectional affix and the other finite forms are stressed on the stem) present a problem for the Learner. It forms the imperatives by the ∅-allomorph, palatalizing the final consonant, instead of adding the *-i* suffix. Representative examples are given in (33).

(33)

	1sg	***2sg***	***3pl***	***Imperative***	***Gen. form***	***Gloss***
a.	d^{j}eržú	d^{j}éržiš	d^{j}éržat	d^{j}erží	*d^{j}érž	'hold'
b.	var^{j}ú	vár^{j}iš	vár^{j}at	var^{j}í	*vár^{j}	'boil'
c.	toropl^{j}ú	toróp^{j}iš	toróp^{j}at	torop^{j}í	*toróp^{j}	'hurry'
d.	šučú	šút^{j}iš	šút^{j}at	šut^{j}í	*šút^{j}	'joke'

Here, the learner assumes that if the 2sg or the 3pl form is stressed on the stem, the imperative will be derived using -∅. This is the major group of verbs whose imperatives are derived incorrectly from the 2sg and 3pl forms of the verb. The mobile stress pattern is relatively rare in Russian and therefore the imperatives are derived incorrectly from the 2sg and the 3pl for only 7% of the verbs.

7 Discussion

In this section I discuss several issues arising from the results of the quantification analysis above.

7.1 Multiple bases

Does a two-base approach resolve the problems encountered in imperative formation, and if so, which two bases are needed? It is expected that considering multiple bases of the verb would produce better results and increase the success rate of deriving any paradigmatic form, and quantificational analysis showed that this is indeed the case. The percentage of correctly predicted imperatives increases significantly when using two bases compared to one base. While

the absolute increase in success rate ranged from 2% to 4% depending on the multiple bases used in comparison to the single base, entire classes of verbs were able to benefit from it.

The use of the 1sg form together with any other form resolved the problem with verbs exhibiting mutation only in the 1st sg form (pattern 1). Considering the 2sg or the 3pl form along with the 1sg form helped to alleviate problems with the class of verbs involving mobile stress, generally problematic for 2sg or 3pl bases. The infinitive clearly benefitted the most from being combined with any of the finite bases as information about the verbal class became available which allowed for the correct choice of the NON-PAST stem, providing access to the information about which suffix (*-a-* or *-aj-*, *-ova-* or *-uj-*) must surface in the imperative.

Lastly, since most problems identified in section 6.1, 6.2, and 6.3 are resolved by considering two forms, we would not expect the combination of all four forms to produce significantly better results than the combinations of two forms. This is exactly what we found: the combination of the four forms fares as well in deriving the imperative as the following combinations: 1sg+2sg, 1sg+3pl, 1sg+inf, 2sg+inf, 3sg+inf.

7.2 Acquisition errors in imperative formation

One of the important questions to answer is whether the Minimal Generalization Learner can account for acquisition errors in the formation of the imperatives, and can predict where errors are likely to be observed. Unfortunately, literature on children's acquisition of verbal paradigms in Russian is unknown to me. However, a Google-search revealed several weblogs/ forums describing children's errors in imperative formation. Examination of the forms surfacing in children's speech provides evidence that here the infinitive is used as a base for imperative formation (34).

(34)

	Infinitive	***Imperative***	***Children's form***	***Gloss***
a.	p^{j}isát^{j}	p^{j}iší	p^{j}isáj	'write'
b.	sosát^{j}	sos^{j}í	sosáj	'suck'
c.	pr^{j}átat^{j}(s^{j}a)	pr^{j}áč(s^{j}a)	pr^{j}átaj(s^{j}a)	'hide (oneself)'
d.	l^{j}izát^{j}	l^{j}iží	l^{j}izáj	'lick'
e.	spát^{j}	sp^{j}í	spáj	'sleep'

f.	r^{j}isovát^{j}	r^{j}isúj	r^{j}isováj	'draw'
g.	tancevát^{j}	tancúj	tancováj	'dance'

No errors of other types were recorded, though given the paucity of the acquisition data, no real conclusions can be drawn. What is interesting, however, is that all of the errors in (34) are the types of errors also produced by the Learner, and they all involve the *-at*^{j} infinitive type. As I showed above, the infinitive does not show a distinction between *-a-* and *-aj-* verbs, and the Learner generates the imperatives of *-a-* verbs as if they belong to the *-aj-* class. Also, the Learner tends not to replace the *-ova-* suffix with *-uj-*. Examples (34a-e) present the cases when children postulate the existence of the suffix *-aj-* for verbs of *-a-* class, and there are similar errors generated by the Learner as in (21). Examples (34f-g) show that children also fail to replace *-ova-* with *-uj-*: these forms are similar to the imperative forms generated by the Learner in (23).

A distinction similar to the one between *-a-* and *-aj-* classes of verbs exists between *-e-* and *-ej-* classes. The infinitival form of the verb does not provide information about which of the suffixes, *-e-* or *-ej-*, appears in the non-past forms. As I showed above, this similarity between *-a-/-aj-* and *-e-/-ej-* verbs does not extend to the predictions about imperative derivation. While for the former class of verbs, the imperatives of *-a-* verbs were incorrectly derived with the verbal suffix *-aj-*, the opposite is true for the latter class: the imperatives of *-e-* verbs are derived correctly, while the imperatives of the *-ej-* verbs are incorrectly assumed to have *-e-* suffix. These results stem from frequency effects: the more frequent the class is, the higher the confidence values assigned to the rules generating the correct imperatives for this class. For this reason, the rule which changes *-at*^{j} to *-aj* in order to form the imperative has a higher confidence value than the competing rule changing *-at*^{j} to *-i*, and is applied to the majority of *-at*^{j} verbs, resulting in incorrect imperatives for *-a-* verbs. On the contrary, the rule which changes *-et*^{j} to *-i* has higher confidence value than the rule changing *-et*^{j} to *-ej*, and the opposite effects are observed. Assuming that the Learner mimics the real-life language learning process, we expect to find similar over-application of rules in children's speech. These expectations were found to be justified: no mistakes for *-e-* class verbs were found on-line.

These preliminary results show that children in fact use the infinitive as a base for the imperative, and the Learner models a stage in language acquisition correctly. The possibility of deriving the imperative from the infinitive correctly for 87% of the verbs makes learners think that the infinitive is a possible base, since it can account for a majority of the imperatives they hear.

7.3 Frequency and markedness of the base

The fact that Russian learners use the infinitive to derive the imperative of some verbs must be justified based on frequencies of the forms of the verbal paradigm in the input children hear. Vakar (1966) presents data about the frequencies of various verbal forms in spoken Russian. However, the frequencies of the individual members of the verbal paradigm are not given there: frequencies of particular persons and numbers are given separately, and not combined. I approximate the frequencies of the individual members of the verbal paradigm by multiplying the frequency of the particular person by the frequency of particular gender by the frequency of the mode (indicative). For instance, to find the frequency of the 1sg non-past form, I use the approximation in (35):

(35) $freq(\text{1sg non-past}) = freq(\text{1st}) \times freq(\text{sg}) \times freq(\text{indicative})$

The results are given in (36):

(36) a. *Frequencies of the individual members of the non-past verbal paradigm (based on Vakar(1966))*

	singular	***plural***
1st	14.30%	4.74%
2nd	22.44%	7.44%
3rd	15.08%	5.00%

b. *Frequency of the infinitive (Vakar 1966):* 15.30%

As one can see from (36), the most frequent non-past verbal form is 2sg, followed by the infinitive, the 3sg, and the 1sg forms. Because of comparatively high frequency of the infinitive form in the spoken Russian, it is unsurprising that learners use it as a base.

This leads to the question of whether the most frequent form of the paradigm always serves as its base (Bybee 2001). Our findings show that for imperatives this categorical statement is invalid. While being the most frequent, the 2sg does not fare any better than other, less frequent forms such as the 3pl and 1sg. And the 3pl, being much less frequent than the 1sg form, is not significantly worse in predicting the imperative. Therefore, high frequency of a verbal form does not necessarily mean than it serves as a better base for deriving the other member of the paradigm.

Similarly, the least marked forms are also not necessarily better bases (cf. Kenstowicz 1996, Benua 1997). If relative markedness of the forms within Russian verbal paradigms is evaluated in terms of consonant mutations, we would expect the 3pl form to be the least marked in the non-past paradigm.

(37) a. Mutation pattern 1: 1sg
b. Mutation pattern 2: all non-past forms
c. Mutation pattern 3: 2sg, 3sg, 1pl, 2pl

There are no mutation patterns where the 3pl form has a mutated consonant while other forms do not (not counting the exceptional pattern 4, occurring with just one verb). Further, if the mutation occurs in the 3pl form, it must occur in all other non-past forms (mutation pattern 2). The 3pl form, however, is not the best base in predicting the imperative, as its success rate in deriving the imperative is statistically the same as the success rate of any other non-past form.

8 Conclusion

In this paper I tried to achieve two main goals. First, I evaluated the status of different verbal forms in Russian as potential predictors for the imperative. Based on my study, the debate in the literature about which form serves as the base for the imperative is justified, as all finite forms perform equally well at this task, and none of them can be clearly chosen as superior. Yet

none of the finite forms is entirely successful as each fails to derive the imperative for a certain class of verbs. Also, as expected, my study showed that the infinitive provides the worst results, failing to derive imperative forms correctly for large classes of verbs, though it turned out to be more successful than anticipated.

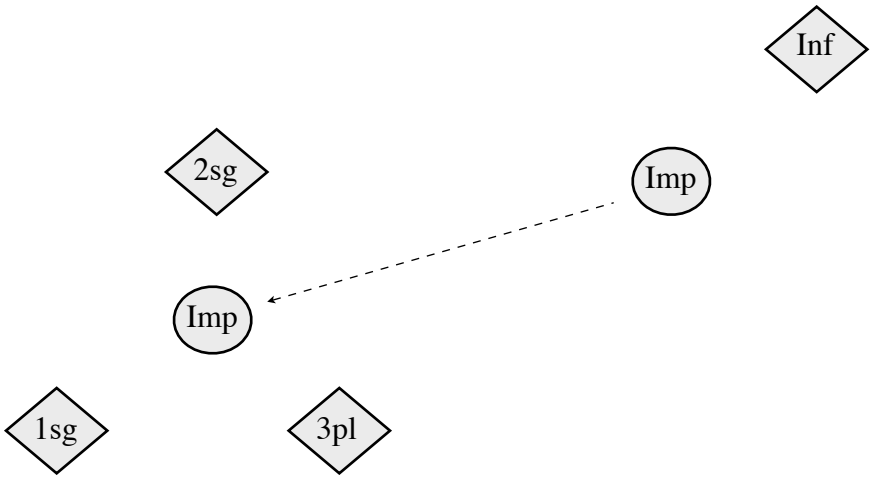
The Albright claim that there is no need to resort to the underlying representation and that a free-standing form can be used as a base for the paradigm was confirmed. While none of the forms of the verbal paradigm achieves 100% accuracy in predicting the imperative, one form is enough to achieve up to 95% success rate. Such a result is initially unpredictable, given the complexity of the Russian verbal paradigm which spans several dimensions, including various stress patterns and segmental changes that are not entirely predictable from a single form.

Further, since all the non-past forms were found to be statistically the same in terms of predicting the imperative, the task of learning the imperative derivation is simplified for language learners as there is no need to select a single base; this selection task is one of the most computationally complex parts of Albright's model of base discovery, and in the case of Russian imperative derivation it can be avoided; also limited input does not present problems for imperative formation: no matter which finite form dominates the input to the learner, the imperatives will be acquired with the same success rate.

This insignificant difference in success rates of predicting the imperative for various forms of the paradigm eliminates the difference between Optimal Paradigms (OP) approach (McCarthy 2005) and Albright's model of base discovery. The imperative form can be treated as a form which is derivable equally well either when one considers the entire paradigm simultaneously under the OP approach, without choosing a base, or when one postulates a certain non-past form as a dedicated base. Equal similarity of the imperative to all non-past forms allows it to be treated as a certain point of equilibrium: connections of it to any other non-past forms are on average statistically same. This property explains the uniqueness of the imperative form in Russian verbal paradigm.

For learners, acquiring Russian, at the initial stage the imperative exhibits the strongest connection with the infinitive. Later, it shifts away from the infinitive and its connection with the infinitive weakens, while the connection with the non-past forms strengthens until all the connections with non-past forms become equal. Schematically this shift is illustrated in (38).

(38)



Second, I proposed an extension to the Minimal Generalization Learner model to deal with multiple bases. I tested this model on Russian imperatives and demonstrated that this extension provides significant improvement over the original, single-base model. I then considered what learnability problems arise when imperatives are derived from a single base. I further examined the errors in imperative formation occurring in the speech of Russian learners, and demonstrated that the predictions made by the Minimal Generalization Learner about the problems in the derivation of imperatives are borne out. Based on the frequency of the infinitive and its fairly high success rate in deriving the imperatives, it is unsurprising that it is often chosen as a base for imperative formation by language learners.

While in this paper I demonstrated that there is no need to rely on underlying representation, such as the basic or non-past stem, in order to form the imperative, it was nevertheless not shown whether the Albrightian approach of deriving the imperative from free-standing forms is superior to Jakobsonian approach of deriving the imperative from the abstract stem. In order to provide such a comparison, one needs to analyze from a quantificational point of view how easy it is to learn the basic stem, and how easy it is to learn the derivation of the imperative from this stem. The former requires running the Minimal Generalization Learner with the basic stem serving as an output, and analyzing the success rate of the basic stem derivations. The latter requires running the Learner with the basic stem used as an input, and the

imperative as an output. Combining these two parts will provide the answer to the question of whether using the abstract basic stem is superior to using free standing forms as a basis for the imperative. I leave this for future research.

References

Albright, Adam, 2002. The Identification of Bases in Morphological Paradigms. Doctoral dissertation, UCLA, Los Angeles, CA.

Albright, Adam. 2005. The morphological basis of paradigm leveling. In *Paradigms in Phonological Theory*, ed. by Laura Downing, Tracy Alan Hall, Renate Raffelsiefen, 17-43. Oxford: Oxford University Press.

Albright, Adam and Bruce Hayes. 1999. An Automated Learner for Phonology and Morphology. ms. Available at http://www.linguistics.ucla.edu/people/hayes/learning/

Albright, Adam and Bruce Hayes. 2004. Modeling Productivity with the Gradual Learning Algorithm: The Problem of Accidentally Exceptionless Generalizations. In *Gradience in Grammar: Generative Perspectives*, ed. by Gisbert Fanselow, Caroline Fery, Matthias Schlesewsky and Ralf Vogel, 185-204. Oxford: Oxford University Press.

Bar-Shalom, Eva and William Snyder. 1999. On the Relationship between Root Infinitives and Imperatives in Early Child Russian. In *BUCLD 23 Proceedings*, 56-67. Somerville, MA: Cascadilla Press.

Benua, Laura. 1997. Transderivational Identity: Phonological Relations Between Words. Doctoral dissertation, University of Massachusetts Amherst.

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

Berger, Tilman. 1981. The formation of the imperative in Modern Russian. *Russian Linguistics* 6:65-80.

Es'kova, N. A. 1985. K morphologii russkogo imperativa (forma vtorogo litsa edinstvennogo chisla). [On the morphology of the Russian imperative (2nd person singular form)]. *Russian Linguistics* 9:149-63.

Gvozdev, A. N. 1961. *Sovremennyj russkij literaturnyj jazyk*. Moscow: Gos. Uč. Ped. Iz

Jakobson, Roman. 1948. Russian Conjugation. *Word* 4:155-67.

Kenstowicz, Michael. 1996. Base-Identity and Uniform Exponence: Alternatives to Cyclicity. In J. Durand & B. Laks, eds., *Current Trends in Phonology: Models and Methods*. Salford: University of Salford, 363-394

McCarthy, John. 2005. Optimal Paradigms. In *Paradigms in Phonological Theory*, ed. by Laura Downing, Tracy Alan Hall, Renate Raffelsiefen, 170-210. Oxford: Oxford University Press.

Mikheev, A. 1997. Automatic Rule Induction for Unknown-Word Guessing. *Computational Linguistics* 23:405–23.

Pinker, Steven and Alan Prince. 1988. On language and connectionism: Analysis of a Parallel Distributed Processing model of language acquisition. *Cognition* 28:73–193.

Pulkina I., and E. Zakhava-Nekrasova. 1974. *Russian: a practical grammar with exercises*. [translated from Russian by V. Korotky; edited by R. Dixon and G. Konorova] Moscow: Russky Yazyk.

Sharoff, Serge. The frequency dictionary for Russian. http://www.comp.leeds.ac.uk/ssharoff/frqlist/frqlist-en.html

Švedova, N. Yu. 1982. *Russkaja Grammatika* [Russian Grammar]. Moscow: Nauka.

Swan, Oscar. 1984. The morphohonemics of the Russian imperative. *Russian Linguistics* 8:39-47.

Timberlake, Alan. 2004. *A Reference Grammar of Russian*. Cambridge: Cambridge University Press.

Townsend, Charles E. 1980. *Russian Word-Formation: corrected reprint*. Columbus: Slavica Publishers.

Vakar, N. P. 1966. *A Word Count of Spoken Russian: The Soviet Usage*. Columbus: The Ohio State University Press.

Vinogradov, V. V. 1972. *Russkij Jazyk*. [Russian Language]. Moscow: Vysshaja shkola.

Zaliznjak, A. A. 1977. *Grammaticheskij slovar' russkogo jazyka*. [Grammatical Dictionary of Russian]. Moscow: Russkij Jazyk.