Learning complex inflectional paradigms through blended gradient inputs

Eric Rosen
Johns Hopkins University
errosen@mail.ubc.ca

Abstract

Through Gradient Symbolic Computation (Smolensky and Goldrick, 2016), in which input forms can consist of gradient blends of more than one phonological realization, we propose a way of deriving surface forms in complex inflectional paradigms that dispenses with direct references to inflectional classes and relies solely on relatively simple blends of input expressions.

In languages whose inflectional systems have a highly complex paradigmatic structure, it becomes a challenge to explain how a speaker can produce a correct inflectional form when the number of possible forms is exceedingly large. As Ackerman and Malouf (2013, p. 429) (henceforth A&M) comment: "That speakers are able to do this is a truly puzzling accomplishment given the extraordinary variation and complexity attested in the morphological systems of the world." When it becomes extremely difficult for a speaker to memorize every possible inflectional form for every lexeme in their language, it becomes attractive to posit some means by which they can produce a correct unmemorized form.

Three interacting types of paradigmatic complexity that put a burden on learning are addressed here: (a) differences in inflectional material across lexemes, which, descriptively, result in division of lexemes into inflectional classes; (b) syncretism, where the same inflectional material occurs in different paradigm cells; (c) independent subsystems of inflectional class behaviour, evident in Mazatec (see §3.) These complexities are all ways in which inflectional patterns depart from *canonicity* as defined in detail by Corbett (2009); Baerman and Corbett (2010); Stump (2016).

Through Gradient Symbolic Computation (Smolensky and Goldrick, 2016), a type of Harmonic Grammar, and with examples from Russian and Mazatec[1], we propose a system that enables a speaker to produce correct forms of complex paradigms through learnable input representations without indexing to inflectional classes.

The paper is organized as follows. §1 introduces GSC and how it can be applied to learning exponents of inflectional paradigms. §2 shows how Russian noun inflection can be acquired through this framework. §2.1 shows how GSC limits the kinds of inflectional patterning that are possible. §3 analyses complex paradigms in Mazatec, where inflectional classes vary in three cross-cutting dimensions. §4 discusses testing and comparison with other models. §5 summarizes how this framework can both explain departures from canonicity in inflectional paradigms and also constrain the degree of departure from the canonical.

1 Gradient Symbolic Computation

We adopt here Gradient Symbolic Computation (henceforth GSC), in which gradient inputs are given numerical activation levels, which we shall show to be learnable, and are evaluated with weighted constraints that calculate the Harmony of input candidates, where the candidate with the greatest Harmony surfaces. This formalism is part of a larger research program in which computation derives outputs from gradient representations in phonology, syntax and semantics (Cho et al., 2017; Faust and Smolensky, 2017; Faust, 2017; Goldrick et al., 2016; Hsu, 2018; Müller, 2017; Rosen, 2016, 2018; Smolensky et al., 2014; Smolensky and Goldrick, 2016; Van Hell et al.,

[1]In response to a reviewer's question about the effects and degree of simplification of the paradigms of these two languages, we contend that (a) any linguistic analysis will have some degree of simplification, (b) the paradigms are no more simplified than those analysed by A&M and in fact include *more* paradigm positions than are represented in their table A6 (p. 457), and (c) nothing in the analysis was excluded because the model couldn't handle it.

Proceedings of the Society for Computation in Linguistics (SCiL) 2019, pages 102-112.
New York City, New York, January 3-6, 2019

2016; Zimmermann, 2017b,a, forthcoming).[2]

In the examples in §2 from Russian noun inflection, the surface exponents that can represent genitive singular among the descriptive inflectional classes, form a blend of input segments, which we shall call 'inflectional input': $\{a, i\}$. A lexeme (basic element of the lexicon) in descriptive class 1 whose genitive sg. is *vina*, 'wine' contributes an input we shall call 'base input' that consists of what is traditionally thought of as a stem plus a blend of word-final segments $\{a, u, o\}$ that mirrors segments that can occur in surface-final position. We propose that it derives as follows, with numerical details shown later in table 4.

	Lexical base	Inflectional affix
Input:	$\overbrace{\text{vin}}^{\text{root}}\{a, u, o\}$	GEN.SG. = $\{a, i\}$
Output:	$\underbrace{\text{vin}}_{\text{root exponent}}\ \underbrace{\text{a}}_{\text{affixal exponent}}$	

Table 1: Gen.sg. of lexeme, descriptive class 1

In this framework, a discrete output form is chosen from possible candidates by the action of Faithfulness constraints that are familiar from Optimality Theory (Prince and Smolensky, 1993), but which, in GSC, have weighted values and evaluate gradient input activations. We assume a highly-weighted *quantization* constraint quantified by an equation in Cho et al. (2017) that disfavours blended outputs and gives a higher Harmony to discrete non-blended forms. The tableaux below thus ignore blended output candidates.[3]

Surface material that some models view as deriving solely from an affix input is derived here from two distinct input sources, thus blending two competing analyses: a constructivist approach that builds whole words out of morphemes and a word-based approach that seeks relationships between whole word forms. The input representation of a lexeme that includes affixal material as an intrinsic part of the word is based upon a whole-word model; an affix as a fully independent input follows a morpheme-based model. A similar approach is taken by Smolensky and Goldrick (2016), treating French liaison as derived from both the end of a preceding word and the beginning of a following word, thus combining two competing approaches to liaison in the literature.

CLASS	1	2	3	4
		SINGULAR		
NOM	-o	-∅	-a	-∅
ACC	-o	-∅	-u	-∅
GEN	-a	-a	-i	-i
DAT	-u	-u	-e	-i
LOC	-e	-e	-e	-i
INST	-om	-om	-oj	-ju
		PLURAL		
NOM	-a	-i	-i	-i
ACC	-a	-i	-i	-i
GEN	-∅	-ov	-∅	-ej
DAT	-am	-am	-am	-am
LOC	-ax	-ax	-ax	-ax
INST	-am'i	-am'i	-am'i	-am'i

Table 2: Russian noun paradigm

Any matching phonological material from the two sources will combine through coalescence. When there are multiple descriptive inflectional classes, the exponent for a given inflectional combination varies, depending not only on the lexeme, but on the stem when there is stem allomorphy (Stump, 2016). The current proposal for blended word-final input segments that mirror affixes, directly encodes the fact that lexeme/stem choice will affect affix-choice. Non-concatenative and suppletive alternations can thus be represented as well, as blends of different stem alternants occurring underlyingly.

2 Learning Russian noun inflection

For ease of exposition, we first examine the relatively simple paradigms of case/number noun suffixes in Russian, with data from A&M p. 460 given in table 2. For presentation purposes, we ignore here the paradigm defectiveness of some Russian nouns (Corbett, 2007).

The exponent that surfaces for a given stem and person/number combination is the one with the highest aggregate activation that surpasses a threshold determined by MAX and DEP constraints to be defined below. When two instances of the same exponent occur in both base and inflectional input, they can coalesce together in the output, with an aggregate input activa-

[2] A reviewer asks what advantages this model has over a "genuine connectionist model" such as Goldsmith and O'Brien (2006); Kann and Schütze (2016); Malouf (2018). In this model, the knowledge it contains is completely transparent, whereas in the last two models cited by the reviewer, it is not clear what kind of knowledge is contained in those networks. (See §4 for further discussion.)

[3] But see Zimmermann (2017b,a, 2018) for analyses of other phenomena in which outputs are also gradient.

Source	☞a	u	o	i
Base input	0.1	0.02	0.16	
GEN-SG	0.3			0.26
Total	**0.4**	0.02	0.16	0.26

Table 3: Summed activations for class 1 GEN-SG

vin$(0.1 \cdot a, 0.02 \cdot u, 0.16 \cdot o)$ 'wine' + GEN-SG$(0.3 \cdot a, 0.26 \cdot i)$			
	MAX 5	DEP -2	Harmony
☞ **vin-a**	2.0	-1.2	**0.8**
vin-u	0.1	-1.96	-1.86
vin-o	0.8	-1.68	-0.88
vin-i	1.3	-1.48	-0.18
vin-e		-2.0	-2.0
vin			0.0

Table 4: Tableau for class 1 GEN-SG

tion that is the sum of the two source activations. Consider again the same genitive singular form *vina* given above in table 1. Suppose that the base input has a blend of three affix-mirroring segments with the following activity levels: vin + (0.1·a, 0.02·u, 0.16·o) and affixal GEN-SING is a blend of segments (0.3·a, 0.26·i). There are then four possible candidate segments that have a non-zero input activation: /-a/, /-u/, /-o/ and /-i/, whose input activations will be the sums of any pair of co-occurring segments that can coalesce in the output as shown in table 3.

The segment *a* surpasses the relevant threshold and surfaces, having the highest aggregate input activation. Table 4 shows a harmonic tableau for this form. In GSC, a MAX constraint (here weighted 5.0) contributes positive Harmony proportionate to the surfacing of underlying activation. A DEP constraint (here weighted at -2.0) contributes negative Harmony for the deficit between input and output activations. We assume that quantization will only allow surface activations of 0 or 1 so the tableau only considers candidates with those values. The positive contribution to Harmony from MAX for the winning candidate is is 5 times the sum of its input activations for /a/ which are 0.1 from the base input plus 0.3 from gen.sg. inflection input. The winning candidate has the highest Harmony.

An affixless candidate incurs no MAX reward or DEP penalty and will be optimal when all other candidates have negative Harmony (table 5.)

vin$(0.1 \cdot a, 0.02 \cdot u, 0.16 \cdot o)$ 'wine' + GEN-PL$(0.28 \cdot ov, 0.22 \cdot ej)$			
	MAX 5	DEP -2	Harmony
vin-a	0.5	-1.8	-1.3
vin-u	0.1	-1.96	-1.86
vin-o	0.8	-1.68	-0.88
vin-ov	1.4	-1.44	-0.04
vin-ej	1.1	-1.56	-0.46
vin-e		-2.0	-2.0
☞ **vin**			**0.0**

Table 5: Tableau for class 1 GEN-PL

The following simple algorithm[4] learned blended base and inflectional input values such that the optimal output candidate is the correct target of learning for all 48 forms. Following Faust and Smolensky (2017, page 2), not just an individual segment but also other structures can have an activity level.

- Initialize all activations at zero.
- Calculate the Harmony for each descriptive stem-class/number/case combination.
- If the wrong affix is predicted, decrease its two input activations if nonzero and increase the activations of the desired affix. Stepsize 0.02.
- Repeat until all 48 paradigm positions are correctly predicted.

Tables 6 and 7 show the learned input values after 72 iterations.

Nonzero inputs for an inflectional combination indicate its set of possible exponents. A co-occurring nonzero base input narrows down the choice of those exponents. For example, the locative singular affix is *-e* for all descriptive classes except class 4, where *-i* surfaces. The activation of 0.12 for *-i* for a class 4 base input allows it to surpass other competitors where it surfaces.

In summary, representing both base and inflectional inputs with a blend of partially activated exponents allows a speaker to encode all the information in a multi-class inflectional paradigm directly on the relevant input forms.

[4] A reviewer suggests that the contribution of this model is theoretical rather than computational and that the learning algorithm is 'fairly trivial.' We see no advantage in a model of human learning that is intentionally complex; the focus of this proposal is to contribute to computationally-based theories of human language rather than to test new computational techniques.

Class	1	2	3	4
a	0.1		0.04	
e			0.12	
u	0.02		0.06	
om		0.04		
ov		0.02		
ax				
am				
am'i				
i			0.1	0.12
oj			0.06	
ju				0.06
o	0.16			
ej				0.08

Table 6: Learned nonzero input values

Cell	NSg	ASg	GSg	DSg	LSg	ISg
a	0.28		0.3			
e				0.26	0.3	
u		0.28		0.3		
om						0.3
ov						
ax						
am						
am'i						
i			0.26	0.22	0.24	
oj						0.26
ju						0.26
o	0.24	0.22				
ej						
Cell	NPl	APl	GPl	DPl	LPl	IPl
a	0.24	0.24				
e						
u						
om						
ov			0.4			
ax					0.3	
am				0.3		
am'i						0.3
i	0.3	0.3				
oj						
ju						
o						
ej			0.22			

Table 7: Learned inflectional input values

2.1 Restricting possible paradigms

The Russian noun paradigm exhibits both syncretism within descriptive inflectional classes and variance of exponents for a given position across inflectional classes. Our model predicts that there are limits to such departures from canonicity. In the dative singular, the affix is *e* in class 3 and *i* in class 4. In the genitive singular, it is *i* for both classes, making *i* syncretic across those two feature combinations for class 4. Our model predicts the impossibility of the pattern in the right-hand half of table 8.

	Real		Impossible	
	Class 3	Class 4	Class 3	Class 4
Gen. Sg.	i	i	i	e
Dat. Sg.	e	i	e	i

Table 8: Crossing diagonals

In the impossible paradigm, lines connecting the two instances of *i* (coloured blue) and connecting the two instances of *e* (coloured red) would cross. Such diagonal crossing can be shown to mathematically imply activation hierarchies that lead to a contradiction (table 9) , where ι represents the activation of *i* and ϵ of *e* for a given (subscripted) feature value or descriptive inflectional class.

$$\iota_{g.sg.} + \iota_3 > \epsilon_{g.sg.} + \epsilon_3 \quad (1)$$
$$\epsilon_{g.sg.} + \epsilon_4 > \iota_{g.sg.} + \iota_4 \quad (2)$$
$$\iota_{dat.sg.} + \iota_4 > \epsilon_{dat.sg.} + \epsilon_4 \quad (3)$$
$$\epsilon_{dat.sg.} + \epsilon_3 > \iota_{dat.sg.} + \iota_3 \quad (4)$$
$$(1) + (2) : \iota_3 + \epsilon_4 > \epsilon_3 + \iota_4 \quad (5)$$
$$(3) + (4) : \iota_4 + \epsilon_3 > \epsilon_4 + \iota_3 \quad (6)$$
$$(6) \text{ contradicts } (5)$$

Table 9: Contradictory inequalities

Thus, in spite of a common misconception that "you can do anything with numbers", the proposed GSC model predicts that many conceivable paradigms are impossible.

3 Learning inputs in Mazatec

We now consider verb paradigms in Chiquihuitlán Mazatec, a language with a high degree of paradigmatic complexity. Jamieson (1982) shows that its verbal inflectional morphology has three separate dimensions in the affirmative forms of the neutral and incompletive aspects: (a) a stem formative (table 12), with 35 possible exponents, (b) a final vowel (table 19), with 11 possible exponents, and (c) a tone pattern (table 15), with 15 possible exponents. For example, in the neutral 1st person singular form in table 10, ba^{3}sæ^1, the stem is merely the segment *s*. The stem formative is *ba*, the final vowel is *æ* and the tone pattern is 3-1. Yet these $35 \times 11 \times 15 = 5775$ hypothetically

	NEUTRAL		INCOMPLETIVE	
	SG	PL	SG	PL
1INCL		ča^{2}sẽ^{2}		ča^{2}sẽ^{42}
1	ba^{3}sæ^{1}	ča^{2}sĩ^{24}	k^{w}a^{3}sæ^{1}	ča^{4}sĩ^{24}
2	ča^{2}se^{2}	ča^{2}sũ^{2}	ča^{4}se^{2}	ča^{4}sũ^{2}
3	ba^{3}se^{2}		k^{w}a^{4}se^{2}	

Table 10: Paradigm for 'remember'

possible combinations are actually reduced to 20 in Jamieson's list of paradigms. Table 10 gives the paradigm for *ba^{3}se^{2}* 'remember' from A&M:450, taken from Jamieson (1982, p. 167). This lexeme is in stem formative descriptive class 11, tone pattern class B 3-1 and final vowel class 2.

Combinations of morphosyntactic features are represented by combinations of multiple exponents, but there is no apparent correspondence between individual feature values and individual exponents. The surface form for a verb expresses feature values for aspect, person and number, with a stem formative, a tone pattern and a suffix vowel.

Stems in many cases consist merely of a single segment. As A&M:447 observe, a stem is much more easily identified by "its membership in an inflectional class in each of three distinct cross-cutting dimensions: tone pattern, final vowel, and stem formative." In the proposed blended representations, a lexeme's base input carries much of the information that predicts what exponents it occurs with, in this way identifying what can descriptively be called its stem. It also directly encodes what is descriptively its inflectional class in a way that is not possible in a model in which paradigm information and stem information are separate. Table 11 shows, with activation values omitted for now, an example basic input-output structure, which contains gradient blends of consonants in the input.

Tone and final vowel omitted here		
	Lexical base	Stem formative
Input:	{b, k^{w}, č} a $\overbrace{\text{s}}^{\text{root}}$	3rd.p.def.incomp. { s͡k, k^{wh}, s, k^{w} }
Output:	$\underbrace{\text{k}^{w}\text{a}}_{\text{stem formative}}\ \underbrace{\text{s}}_{\text{root exponent}}$	

Table 11: Input-output for 'remember' 3rd.p.def.inc.

Some further complexities in Jamieson's data are omitted here for simplicity of exposition but can be explained by our analysis. These include suffixing of verbs, verb compounding, verbs with no stem formative[5] and stem suppletion.[6]

Class	Stem formatives				Representations	
	3def or 1s		Oth.prsns			
	Neu.	Incomp.	Neu.	Incomp.	blend	vowel
1	be	k^{w}e	be	k^{w}e	b,**k^{w}**	e
2	ba	k^{w}a	ba	k^{w}a	b, **k^{w}**	a
3	bo	s͡ko	čo	čo	b, **s͡k**, č	o
4	bu	s͡ku	ču	ču	b, **s͡k**, č	u
5	hu	s͡ku	č^{h}u	č^{h}u	h, **s͡k**, č^{h}	u
6	hi	s͡ki	č^{h}i	č^{h}i	h, **s͡k**, č^{h}	i
7	^{h}ba	k^{wh}a	^{h}ba	k^{wh}a	^{h}b, **k^{wh}**	a
8	¢i	¢i	nĩ	nĩ	**¢**, n	i
9	su	su	nũ	nũ	**s**, n	u
10	bu	ku	bu	ku	b,**k**	u
11	ba	k^{w}a	ča	ča	b, k^{w}, č	a
12	ka	s͡ka	ča	ča	k, **s͡k**, č	a
13	^{h}ba	k^{wh}a	nã	nã	^{h}b, **k^{wh}**, n	a
14	ba	k^{w}a	nã	nã	b,**k^{w}**, n	a
15	bi	k^{w}i	bi	k^{w}i	b,**k^{w}**	i
16	bu	s͡ku	^{n}tu	^{n}tu	b, **s͡k**, ^{n}t	u
17	hi	si	či	ši	h, s, č, š	i
18	^{h}ba	k^{wh}a	č^{h}a	č^{h}a	^{h}b, **k^{wh}**, č^{h}	a

Table 12: Representations of stem formatives

Among the 35 possible stem formatives, 11 possible final vowels and 15 possible tone patterns, a given lexeme's input form only needs a blend of a small subset of each of these exponents.

3.1 Learning inputs for stem formatives

Consider first the stem formatives, shown in table 12 (Data from Jamieson (1982).)[7]

The proposed representation for each base input is simply a blend of all its possible initial consonants. The vowels stay the same across a class. Which consonant surfaces depends on how this blend interacts through coalescence of identical consonants with an inflectional input blend of consonant features.

Tables 13 and 14 show the results of the same kind of algorithm discussed on page 3 for learning input activations of inflectional person-number combinations and base inputs respectively.

It took 22 iterations to learn activations that derive the correct stem formative consonants for each of 4 person-number groups in 18 classes.

Abstracting away for now from tone and final

[5]If a stem formative is absent from the base input, an anchoring constraint will prevent an inflection-based stem-formative input from surfacing.

[6]Stem suppletion can be accounted for by an input that includes a blend of different stem allomorphs each linked to different inflectional features.

[7]Following Golston and Kehrein (1998), we take what Jamieson transcribes as complex onsets and nuclei to be simplex segments with secondary articulations.

3def,1sg,neu	
b	0.27
k	0.06
h	0.12
^{h}b	0.12
Other Neu.	
b	0.21
č	0.21
n	0.21
^{n}t	0.12

3def,1sg,incomp	
$\widehat{sk}$	0.21
k^{wh}	0.21
s	0.12
k^{w}	0.12
Other Incomp	
č	0.15
k^{wh}	0.12
n	0.21
š	0.15

Table 13: Input activations for person/number

vowels, for $k^{w}a^{4}se^{2}$ 'remember' (class 11), 3rd p. def. incompl., the activations for the input given above in table 11 are as follows.

Base input: (0.1·b, 0.3·k^{w}, 0.3·č)as

Inflectional input:
(0.21·$\widehat{sk}$, 0.21·k^{wh}, 0.12·s, 0.12·k^{w})

For each candidate, an inflectional input can coalesce with a base input. The segment with the highest combined activation surfaces, namely k^{w}, with an underlying activation of $0.3+0.12 = 0.42$. We assume the stem consonant and the vowel of the stem formative to both have full 1.0 underlying activations; thus no Harmony is gained by having a gradient inflectional stem formative consonant incorrectly coalesce with the stem consonant nor for a gradient inflectional final vowel to incorrectly coalesce with the stem formative vowel.

3.2 Learning inputs for tone patterns

Tone patterns, which fall into 11 descriptive classes, were also learned by an algorithm that separately learned patterns for the first and second syllables.

Table 15 shows tone patterns listed by class and person-number-aspect combination. (Data from Jamieson (1982).)

There is some phonological predictability for the tones of the second syllable, where 1 is the highest tone and 4 the lowest:

- The initial tone on the second syllable can be no lower than the final tone of the first syllable.
- If the initial tone of the second syllable is 3 or 4 and the same as the final tone of the first, there is a further rise of two tone levels:

– e.g. 4 can be followed by 42 but not simple 4;
– 3 can be followed by 31 but not 3.
– 2 can only be followed by 2 or 24.
– 1 can only be followed by 1.

- We can take the level 4 tone that occurs last in

Class	stem formative	input activation
1	b	0.2
	k^{w}	0.3
2	b	0.1
	k^{w}	0.3
3	b	0.1
	$\widehat{sk}$	0.2
	č	0.2
4	b	0.1
	$\widehat{sk}$	0.2
	č	0.3
5	$\widehat{sk}$	0.2
	h	0.2
	č	0.3
6	$\widehat{sk}$	0.2
	h	0.2
	č	0.3
7	^{h}b	0.3
	k^{wh}	0.2
8	c̸	0.4
	n	0.2
9	n	0.2
	s	0.3
10	b	0.2
	k	0.3
Class	stem formative	input activation
11	b	0.1
	k^{w}	0.3
	č	0.3
12	$\widehat{sk}$	0.2
	k	0.3
	č	0.3
13	^{h}b	0.2
	k^{wh}	0.2
	n	0.2
14	b	0.1
	k^{wh}	0.2
	n	0.3
15	b	0.2
	k^{w}	0.2
16	b	0.1
	$\widehat{sk}$	0.2
	^{n}t	0.3
17	s	0.2
	h	0.2
	š	0.3
	č	0.1
18	^{h}b	0.2
	k^{wh}	0.1
	č	0.3

Table 14: Input activations for stem formatives

Neutral aspect						
Class	3def	1s	1in	2s	1ex	2p
A	3-1	3-1	3-31	3-1	3-14	3-1
B1-1	2-2	1-1	2-2	2-2	2-24	2-2
B3-1	3-2	3-1	2-2	2-2	2-24	2-2
C	3-24	14-3	14-42	14-3	14-34	14-3
D1-1	1-1	1-1	3-2	3-2	3-24	3-2
D3-1	3-2	3-1	3-2	3-2	3-24	3-2
Incompletive aspect						
A(1-7)	4-2	3-1	3-31	3-1	3-14	3-1
A(8-18)	4-2	3-1	4-31	4-1	4-14	4-1
B1-1 (1-7)	4-2	1-1	2-2	2-2	2-24	2-2
B1-1 (8-13)	4-2	1-1	4-42	4-2	4-24	4-2
B1-1 (14-18)	4-1	1-1	4-42	4-3	4-34	4-3
B3-1 (1-7)	4-2	3-1	2-2	2-2	2-24	2-2
B3-1 (8-13)	4-2	3-1	4-42	4-2	4-24	4-2
B3-1 (14-18)	4-1	3-1	4-42	4-3	4-34	4-3
C	3-24	14-3	14-42	14-3	14-34	14-3
D1-1 (8-13)	4-2	1-1	4-42	4-2	4-24	4-2

Table 15: Tone patterns

the 1st exclusive to be a separate tonal input for 1st-exclusive.

Tables 16 and 17 show the results of learning algorithms for first and second syllable tones. Ten iterations were required for syllable 1 and 22 iterations for syllable 2, with a stepsize of 0.05 for syllable 1 and 0.1 for syllable 2. Activations were initialized at 0.2 for each base-input and tonal person-number input for each tone pattern that actually occurs. We assume there exists a Harmonic Grammar analysis of the above constraints, in which they have higher weight than Faithfulness, which will otherwise derive the pattern with the highest aggregate activation.

A two-syllable base input has a tone linked to the mora of each syllable. The inflectional input has a linearly ordered sequence of two tones on the tonal tier.[8] A strongly-weighted anchoring constraint will require the tones on the inflectional input to line up with the tones on the base input when they coalesce.

Consider the second person singular incompletive of 'remember', *ča*^{4}*se*^{2}. This is tonal class B-3-1 (8-13) with inputs shown in table 18. When the two input sources coalesce on each syllable, tone level 4 has the highest activation for syllable 1 of 0.25 + 0.1 = 0.35 and tone level 2 on syllable 2 of 0.4 + 0.0 = 0.4., both the correct tones.

3.3 Learning inputs for final vowels

Table 19, with data from Jamieson (1982) shows the 10 descriptive classes of final vowels.

[8] In the 1st person exclusive, a level 4 tone will occur further to the right of other tones.

First syllable tones					
Tone class	1	2	3	4	14
A1-7			0.25	0.15	
A8-18			0.2	0.25	
B1-1-1-7	0.15	0.25		0.05	
B1-1-18-13	0.2	0.25		0.2	
B1-1-14-18	0.15	0.25		0.2	
B3-1-1-7		0.3	0.2	0.1	
B3-1-8-13		0.25	0.15	0.25	
B3-1-14-18		0.25	0.1	0.25	
C			0.2		0.35
D1-1	0.2		0.15	0.25	
D3-1	0.15		0.2	0.25	
Person/number	1	2	3	4	14
3def Neu.		0.05	0.25		
1s Neu.	0.15		0.15		0.1
1in Neu.		0.2	0.15		0.05
2s Neu.		0.15	0.15		0.05
1ex Neu.		0.2	0.15		0.05
2p Neu.		0.1	0.1		
3def Inc.			0.15	0.35	
1s Inc.	0.25		0.15		0.1
1in Inc.		0.05	0.05	0.1	0.05
2s Inc.		0.05	0.1	0.1	
1ex Inc.		0.05	0.05	0.1	
2p Inc.		0.05	0.05	0.1	

Table 16: Input activations for σ_1 tones

Second syllable tones					
Tone class	1	2	3	4	24
A1-7	0.2	0.2	-0.1		
A8-18	0.5		0.2	0.2	
B1-1-1-7	0.2	0.2			
B1-1-18-13		0.4		0.2	
B1-1-14-18	0.1		0.6	0.4	
B3-1-1-7	0.2	0.2			
B3-1-8-13		0.4		0.2	
B3-1-14-18	0.5		0.4	0.2	
C			0.3	0.3	0.2
D1-1		0.5		0.2	
D3-1	0.1	0.5		0.2	
Person/number	1	2	3	4	24
3def Neu.	0.2	0.4			0.2
1s Neu.	0.5		0.2		
1in Neu.		0.1	0.5	0.8	
2s Neu.	0.2	0.1	0.2		
1ex Neu.	0.2	0.1	0.2		
2p Neu.	0.2	0.1	0.2		
3def Inc.	0.2	0.3	0.1		0.2
1s Inc.	0.5		0.3		
1in Inc.			0.3	0.7	
2s Inc.	0.1		0.3		
1ex Inc.	0.1		0.2		
2p Inc.	0.1		0.2		

Table 17: Input activations for σ_2 tones

The algorithm learns input values, shown in table 20, for $\pm hi$, $\pm lo$, $\pm bk$ and *nas* features, with coalescence of values from the two input sources. A negative value in the table represents a positive value for the negative binary feature: e.g. -0.1

	Syllable 1			
Tone →	1	2	3	4
Tonal class B-31 (8-13)		0.25	0.15	0.25
2nd sg. incomp.		0.05	0.1	0.1
Total		0.3	0.25	**0.35**
	Syllable 2			
Tone →	1	2	3	4
Tonal class B-31 (8-13)		0.4		0.2
2nd sg. incomp.	0.1		0.3	
Total	0.1	**0.4**	0.3	0.2

Table 18: Aggregate activations

Class	3def (basic)	1s	1in	2s	1ex	2p
1	i	æ	ẽ	i	ĩ	ũ
2	e	æ	ẽ	e	ĩ	ũ
3	æ	æ	ẽ	e	ĩ	ũ
4	u	u	ũ	i	ĩ	ũ
5	o	o	õ	e	ĩ	ũ
6	a	a	ã	e	ĩ	ũ
7	ĩ	ẽ	ẽ	ĩ	ĩ	ũ
8	ẽ	ẽ	ẽ	ĩ	ĩ	ũ
9	ũ	ũ	ũ	ĩ	ĩ	ũ
10	ã	ã	ã	ĩ	ĩ	ũ

Table 19: Classes of final vowels

hi means 0.1 $-hi$. We assume that $\forall i$ both $+f_i$ and $-f_i$ do not occur in the same source. Table 21 shows input vowel features for *ča*^{4}*se*^{2} 'remember', final-vowel class 2 in the 2nd sg. incomp. (W_v denotes base input.)

	hi	lo	bk	nas
3def	-0.1			
1s	-0.2	0.2		
1in	-0.2			0.2
2s	0.1	-0.3	-0.2	
1ex	0.4	-0.2	-0.2	0.2
2p	0.4	-0.2	0.3	0.2
	hi	lo	bk	nas
Class 1	0.2		-0.1	
Class 2	-0.2	-0.1	-0.1	
Class 3	-0.2	0.2	-0.1	
Class 4	0.4	-0.2	0.1	
Class 5	-0.2	-0.3	0.1	
Class 6	-0.2	0.1	0.1	
Class 7	0.2	-0.1	-0.2	0.2
Class 8			-0.1	0.2
Class 9	0.4	-0.2	0.1	0.1
Class 10			0.1	0.1

Table 20: Input activations for vowel features

We adopt a second, strongly-weighted quantization constraint from Cho et al. (2017) that requires the sum of activations from a binary feature group such as $\{+hi, -hi\}$ to be 1. Thus $-hi$ competes with $+hi$ and $-hi$ surfaces because its higher aggregate activation results in greater Harmony.

Feature	2nd-sg.	W_v	Winner
hi	0.1	-0.2	$-hi$
lo	-0.3	-0.1	$-lo$
bk	-0.2	-0.1	$-bk$
nas			$0\ nas$

Table 21: Values for final-vowel class 2, 2nd-sg.

$[-hi, -lo, -bk, 0\ nas] = e$, the correct final vowel, is the output after quantization.

4 Testing and comparison with other models

To test how a learner could predict an unseen inflectional form for a given stem, we ran cross-vaidation on Mazatec stem formatives, tone patterns and final vowels, with training set items (70% of the total stemclass×inflection combinations) picked randomly from a Zipf-Mandelbrot distribution. On ten runs for each, the average test accuracy was 83% for stem formatives, 87% on tone 1, 92% on tone 2 and 89% on the final vowel. As a baseline, we simultaneously tested prediction of unseen forms based on frequency of occurring stem formatives, for which the accuracy was 9.5%. The relative success of the model in cross-validation is due to the syncretism that occurs both within paradigms and across classes. The following tableau shows how the correct stem formative k^{wh} is predicted for a never-encountered 3rd-definite-incompletive form of a class 7 stem 'weave' *aʔy*[9] in the holdout set, using activation values that were learned for encountered forms.[10]

The strength of this model is that it directly encodes knowledge of the tendency of an exponent to occur for a given stem and for a given morphosyntactic combination. For example, a Mazatec class 1 stem with a stem-formative blend (0.2·b, 0.3·k^w) encodes the relative tendencies of these two exponents to occur with this stem, with zero-valued exponents having no inclination to occur. In Kann and Schütze (2016) and Malouf

[9]Since the vowel *a* of the stem formative does not vary, for convenience we can take it here to be part of the stem.

[10]Because the activation values were learned just for a random training set, they will be different from those given in tables 13 and 14 above.

3RD-DEF-INC(0.1·a, 0.15·$\widehat{\text{sk}}$, 0.09·k^{wh}, 0.12·s, 0.12·k^{w}) +(0.20·^{h}b, 0.20·k^{wh}) *aʔy-* 'weave'			
	MAX 5	DEP −1	Harmony
$\widehat{\text{sk}}$-aʔy-	0.75	−0.85	−0.10
☞ k^{wh}-aʔy-	1.45	−0.71	0.74
s-aʔy-	0.6	−0.88	−0.28
k^{w}-aʔy-	0.6	−0.88	−0.28
^{h}b-aʔy-	1.0	−0.80	−0.20

Table 22: Tableau for 'weave' class 7 3rd-def-inc.

(2018), there is no indication of what kind of knowledge is represented by a given node or connection. Goldsmith and O'Brien (2006)[11] is more similar to the present model but it is not clear how weights from input to hidden-layer nodes translate into feature-values on a symbolic level. And its winner-take-all mechanism with no threshold to surpass means than zero exponents in a paradigm must be represented as such, rather than resulting from any exponent failing to surpass a threshold. This results in awkward representations when a form is expressed by multiple zero affixes. The advantage of the GSC model, (see Smolensky et al. (2014, p. 1103)) is that it combines a subsymbolic neural level with a symbolic level, i.e. 'microlevel representations and algorithms' with symbolic grammatical theory, with an interface between the two. The subsymbolic level provides a platform for optimization and allows gradient activations. When output activations are quantized to discrete values they percolate to the symbolic level which contains symbolic descriptions that are familiar from symbol-based linguistic theory.

5 Summary

Inflectional patterns in Russian and Mazatec depart from canonicity in a number of ways, as was outlined on page 1. If these systems were completely canonical, each morphosyntactic combination of feature values would have a unique, lexeme-invariant exponent. In our model, in Russian, the variance of an exponent across inflectional classes is expressed directly by a *blend* of segments in an underlying form, e.g. (*a*, *i*) for genitive singular. Syncretism across inflectional feature combinations, such as multiple instances of *-i* in several classes is expressed by its occurrence as a base input, so that it can coalesce with a matching inflectional input[12] for several feature combinations. If a different exponent occurred in each cell and there were no inflectional class divisions, the URs could simply be a single non-blended exponent for each inflectional input, with no inflectional formatives on base inputs. So the representations proposed here directly encode and capture the kinds of departures from canonicity that occur.

In Mazatec, blended inputs capture not just deviation from paradigm canonicity but patterns of predicability within its paradigm structure. Out of 12^{35} possible stem-formative combinations that could occur in 12 paradigm positions, Jamieson only lists 18 classes. Blended input representations, in which no more than four phonemes occur for a base input or inflectional input, limit how many stem formatives occur for a given lexeme and along with the occurrence of the same blended inflectional inputs in multiple person-number combinations, derive the syncretism that occurs across those paradigm positions.

This approach can also derive subtractive morphology, suppletion, umlaut and moraic augmentation. For example, if an inflectional form is subtractive relative to a base, the subtractable part can have partial or negative-valued underlying activation in the UR and express morphosyntactic features that affect its surfacing. Umlaut (Trommer, 2017) can be derived from a blend of different vowel features where alternation occurs.

Given the vast number of ways that languages can deviate from canonical inflection (Corbett, 2009, 2007; Baerman and Corbett, 2010; Stump, 2016), we hope with future research to explore in the GSC framework how it is possible to learn complex paradigmatic patterns in other languages.

6 Acknowledgements

Thanks to Paul Smolensky, Matt Goldrick, Farrell Ackerman, three anonymous reviewers and members of the Johns Hopkins Neurosymbolic Computation Group and the Surrey Morphology Group for valuable discussion and suggestions. Research was generously funded by NSF INSPIRE grant BCS-1344269. All errors are my own.

[11] See also Goldsmith and Rosen (2017), which has similarities with the present model but differs in crucial ways.

[12] In terms of correspondence, this model employs the converse of what Pulleyblank (2008) proposes for his account of reduplication in which a single input has mutiple outputs. Here, pairs of identical input features or segments can each coalesce on a single output feature or segment.

References

Ackerman, Farrell, and Robert Malouf. 2013. Morphological Organization: the Low Conditional Entropy Conjecture. *Language* 89:429–464.

Baerman, Matthew, and Greville Corbett. 2010. A typology of inflectional class interaction. In *14th International Morphology Meeting*. Budapest, Hungary.

Cho, Pyeong Whan, Matthew Goldrick, and Paul Smolensky. 2017. Incremental parsing in a continuous dynamical system: sentence processing in Gradient Symbolic Computation: Supplementary Materials 3: processing. *Linguistics Vanguard* 3.

Corbett, Greville. 2007. Canonical typology, suppletion and possible words. *Language* 83:8–42.

Corbett, Greville. 2009. Canonical inflectional classes. In *Selected Proceedings of the 6th Décembrettes: Morphology in Bordeaux*, 1–11.

Faust, Noam. 2017. How much for that vowel? Talk, Strength in Grammar Workshop, University of Leipzig.

Faust, Noam, and Paul Smolensky. 2017. Activity as an alternative to autosegmental association. In *Manchester Phonology Meeting*. Manchester, United Kingdom.

Goldrick, Matthew, Michael Putnam, and Lara Schwarz. 2016. Coactivation in bilingual grammars: A computational account of code mixing. *Bilingualism: Language and Cognition* 19:857–876. ROA 1441.

Goldsmith, John, and Jeremy O'Brien. 2006. Learning inflectional classes. *Language Learning and Development* 2:219–250.

Goldsmith, John, and Eric Rosen. 2017. Geometrical morphology. *CoRR* abs/1703.04481. URL http://arxiv.org/abs/1703.04481.

Golston, Chris, and Wolfgang Kehrein. 1998. Mazatec Onsets and Nuclei. *International Journal of American Linguistics* 64:311–337.

van Hell, Janet G., Clara Cohen, and Sarah Grey. 2016. Testing tolerance for lexically-specific factors in gradient symbolic computation. *Bilingualism: Language and Cognition* 19:897–899.

Hsu, Brian. 2018. Scalar constraints and gradient symbolic representations generate exceptional prosodification effects without exceptional prosody. Handout, West Coast Conference on Formal Linguistics 36.

Jamieson, Carol. 1982. Conflated subsystems marking person and aspect in Chiquihuitlán Mazatec verbs. *Language* 48:139–167.

Kann, Katharina, and Hinrich Schütze. 2016. Single-Model Encoder-Decoder with Explicit Morphological Representation for Reinflection. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 555–560. Association for Computational Linguistics.

Malouf, Robert. 2018. Generating morphological paradigms with a recurrent neural network. *San Diego Linguistic Papers* 6.

Müller, Gereon. 2017. Gradient symbolic representations in syntax. Handout, IGRA 02: Syntax II.

Prince, Alan, and Paul Smolensky. 1993. Optimality Theory: Constraint Interaction in Generative Grammar. Technical report, Rutgers University Center for Cognitive Science and Computer Science Department, University of Colorado at Boulder.

Pulleyblank, Douglas. 2008. *The nature of the word: Studies in honor of paul kiparsky*, chapter Patterns of Reduplication in Yoruba. MIT Press.

Rosen, Eric. 2016. Predicting the unpredictable: Capturing the apparent semi-regularity of rendaku voicing in japanese through gradient symbolic computation. In *Proceedings of the Berkeley Linguistics Society*, volume 42. ROA 1299.

Rosen, Eric. 2018. Predicting semi-regular patterns in morphologically complex words. *Linguistics Vanguard* 4. ROA 1339.

Smolensky, Paul, and Matt Goldrick. 2016. Gradient Symbolic Representations in Grammar: The case of French Liaison. Rutgers Optimality Archive 1552, Rutgers University.

Smolensky, Paul, Matthew Goldrick, and Donald Mathis. 2014. Optimization and Quantization in Gradient Symbol Systems: A Framework for Integrating the Continuous and the Discrete in Cognition. *Cognitive Science* 38:1102–1138.

Stump, Gregory. 2016. *Inflectional paradigms: Content and from at the syntax-morphology interface*, volume 149 of *Cambridge Studies in Linguistics*. Cambridge, UK: Cambridge University Press.

Trommer, Jochen. 2017. Scalar Cumulativity in German Umlaut. Talk, Strength in Grammar Workshop, University of Leipzig.

Zimmermann, Eva. 2017a. Being (slightly) stronger: Lexical stress in Moses Columbian Salish. Handout, Strength in Grammar Workshop, University of Leipzig.

Zimmermann, Eva. 2017b. Gradient symbols and gradient markedness: A case study from Mixtec tones. Handout: Manchester Phonology Meeting.

Zimmermann, Eva. 2018. Gradient symbolic representations and the typology of ghost segments: An argument from gradient markedness. Handout and slides, Annual Meeting on Phonology, UC San Diego.

Zimmermann, Eva. forthcoming. Gradient Symbolic Representations in the output: A case study from Moses Columbian Salishan stress. In *Proceedings of NELS*, volume 48.