Adaptive analogy in Word-and-Paradigm morphology: the case of Seri verbs

Helen Sims-Williams,[1] Jérémy Pasquereau[2,3] & Matthew Baerman[2]

[1] Centre for Language Evolution, University of Edinburgh [2] Surrey Morphology Group, University of Surrey, [3] University of Poitiers

Funded by the Arts & Humanities Research Council under grant AH/P002471/1 ('Seri verbs'), and the European Research Council ('The evolution of linguistic complexity') under grant 681942. Their support is gratefully acknowledged. Thanks to all our Seri consultants, especially: Deborah Perales, Gabriel Hoeffer, Teresa Hoeffer, Karelia Perales, Genoveva Herrera, and Anamaria Morales.

The language

- Seri is a language isolate spoken on the coast of Sonora (Mexico) in two villages: El Desemboque/Haxöl Iihom and Punta Chueca/Socaaix, by approximately 900 people.

The problem

- Seri verbs have suffixal marking of subject number (singular~plural) and event number (neutral~multiple).
- Considerable allomorphy, alongside paradigmatically disjunctive distribution of allomorphs:

‘hurry’	SG	PL
NEUT	**itanamj**	**itanaml-coj**
MULT	**itanaml-c**	**itanaml-cam**

‘roll’	SG	PL
NEUT	**tmaasij**	**tmaasil-c**
MULT	**tmaasil-im**	**tmaasil-coj**

The problem

- Seri verbs have suffixal marking of subject number (singular~plural) and event number (neutral~multiple).
- Considerable allomorphy, alongside paradigmatically disjunctive distribution of allomorphs:

‘hurry’	SG	PL
NEUT	**itanamj**	**itanaml-coj**
MULT	**itanaml-c**	**itanaml-cam**

‘roll’	SG	PL
NEUT	**tmaasij**	**tmaasil-c**
MULT	**tmaasil-im**	**tmaasil-coj**

- We can make sense of this distribution if we view it as a plurality cline: e.g. *–coj* is always more plural than *–c*.

less plural ⟶ *more plural*

SG NEUT	SG MULT	PL NEUT	PL MULT	
itanamj	**itanaml-c**	**itanaml-coj**	**itanaml-cam**	‘hurry’
tmaasij	**tmaasil-im**	**tmaasil-c**	**tmaasil-coj**	‘roll’

The problem

- Schematically, the distribution looks something like this: any suffix can appear anywhere in the paradigm, but each one is predictably 'right' vs 'left' with respect to any other suffix:

SG NEUT	SG MULT	PL NEUT	PL MULT	
a	**b**	**w**	**x**	lexeme 1
b	**w**	**x**	**z**	lexeme 2

- That means there are systematic relationships between forms in the paradigm, but these are not tied to specific morphosyntactic values.
- This is clearly a problem for a morphemic conception of morphology – but it is also a problem for possible alternatives. E.g. Word-and-Paradigm depends on analogical proportions, but:

SG NEUT		SG MULT		SG MULT		?
b	is to	**w**	as	**b**	is to	**w**

Proposal

- We implement a version of analogical cell-filling that accesses the cells as relative positions on a scale, rather than sets of morphosyntactic features.

SG NEUT		SG MULT		SG MULT		?
b	is to	**w**	as	**b**	is to	**w**

Proposal

- We implement a version of analogical cell-filling that accesses the cells as relative positions on a scale, rather than sets of morphosyntactic features.

~~SG NEUT~~ 0		~~SG MULT~~ 1		~~SG MULT~~ 1		2 (PL MULT)
b	is to	**w**	as	**b**	is to	**w**

Simulation experiments

- We devised a set of computational simulations to demonstrate the effects of these alternative models of production on a Seri-like morphological system.

	NEUT SG (0)	MULT SG (1)	NEUT PL (2)	MULT PL (3)
lift (148)	**b**	**h**	**i**	**k**
bent (333)	**b**	**c**	**g**	**h**
bounce (185)	**r**	**u**	**x**	**z**
cut (111)	**m**	**n**	**w**	**x**
go to bed (37)	**c**	**h**	**w**	**z**

Simulation experiments

- We devised a set of computational simulations to demonstrate the effects of these alternative models of production on a Seri-like morphological system.

	NEUT SG (0)	MULT SG (1)	NEUT PL (2)	MULT PL (3)
lift (148)	**b**	**h**	**i**	**k**
bent (333)	**b**	**c**		**h**
bounce (185)	**r**	**u**	**x**	**z**
cut (111)	**m**	**n**	**w**	**x**
go to bed (37)	**c**	**h**	**w**	**z**

- A random form is deleted from a table of inflectional paradigms.

Simulation experiments

- We devised a set of computational simulations to demonstrate the effects of these alternative models of production on a Seri-like morphological system.

	NEUT SG (0)	MULT SG (1)	NEUT PL (2)	MULT PL (3)
lift (148)	**b**	**h**	**i**	**k**
bent (333)	**b**	**c**		**h**
bounce (185)	**r**	**u**	**x**	**z**
cut (111)	**m**	**n**	**w**	**x**
go to bed (37)	**c**	**h**	**w**	**z**

- A random form is deleted from a table of inflectional paradigms.
- The program tries to predict the form back, using one of three methods:

Simulation experiments

- We devised a set of computational simulations to demonstrate the effects of these alternative models of production on a Seri-like morphological system.

	NEUT SG (0)	MULT SG (1)	NEUT PL (2)	MULT PL (3)
lift (148)	**b**	**h**	**i**	**k**
bent (333)	**b**	**c**	**w**	**h**
bounce (185)	**r**	**u**	**x**	**z**
cut (111)	**m**	**n**	**w**	**x**
go to bed (37)	**c**	**h**	**w**	**z**

- A random form is deleted from a table of inflectional paradigms.
- The program tries to predict the form back, using one of three methods:
    1. **Baseline method**: copy the form used by another verb

Simulation experiments

- We devised a set of computational simulations to demonstrate the effects of these alternative models of production on a Seri-like morphological system.

	NEUT SG (0)	MULT SG (1)	NEUT PL (2)	MULT PL (3)
lift (148)	**b**	**h**	**i**	**k**
bent (333)	**b**	**c**	**c**	**h**
bounce (185)	**r**	**u**	**x**	**z**
cut (111)	**m**	**n**	**w**	**x**
go to bed (37)	**c**	**h**	**w**	**z**

- A random form is deleted from a table of inflectional paradigms.
- The program tries to predict the form back, using one of three methods:
    1. **Baseline method**: copy the form used by another verb
    2. **Set-theoretic analogy**: construct and solve a set-theoretic analogy

MULT SG		NEUT SG		MULT PL		NEUT PL
h	*is to*	**c**	*as*	**h**	*is to*	**c**

Simulation experiments

- We devised a set of computational simulations to demonstrate the effects of these alternative models of production on a Seri-like morphological system.

	NEUT SG (0)	MULT SG (1)	NEUT PL (2)	MULT PL (3)
lift (148)	**b**	**h**	**i**	**k**
bent (333)	**b**	**c**	**h**	**h**
bounce (185)	**r**	**u**	**x**	**z**
cut (111)	**m**	**n**	**w**	**x**
go to bed (37)	**c**	**h**	**w**	**z**

- A random form is deleted from a table of inflectional paradigms.
- The program tries to predict the form back, using one of three methods:
  1. **Baseline method**: copy the form used by another verb
  2. **Set-theoretic analogy**: construct and solve a set-theoretic analogy
  3. **Numeric analogy**: construct and solve an analogy using numeric features.

0		1		1		2
c	*is to*	**h**	*as*	**c**	*is to*	**h**

Simulation experiments

- We devised a set of computational simulations to demonstrate the effects of these alternative models of production on a Seri-like morphological system.

	NEUT SG (0)	MULT SG (1)	NEUT PL (2)	MULT PL (3)
lift (148)	**b**	**n**	***w***	***g***
bent (333)	***r***	***c***	***a***	***x***
bounce (185)	**r**	**s**	**x**	**z**
cut (111)	**a**	**n**	***w***	***g***
go to bed (37)	**a**	**b**	**x**	**y**

- A random form is deleted from a table of inflectional paradigms.
- The program tries to predict the form back, using one of three methods:
  1. **Baseline method**: copy the form used by another verb
  2. **Set-theoretic analogy**: construct and solve a set-theoretic analogy
  3. **Numeric analogy**: construct and solve an analogy using numeric features.
- This is repeated many times, cumulatively altering the system and potentially creating violations of the scale

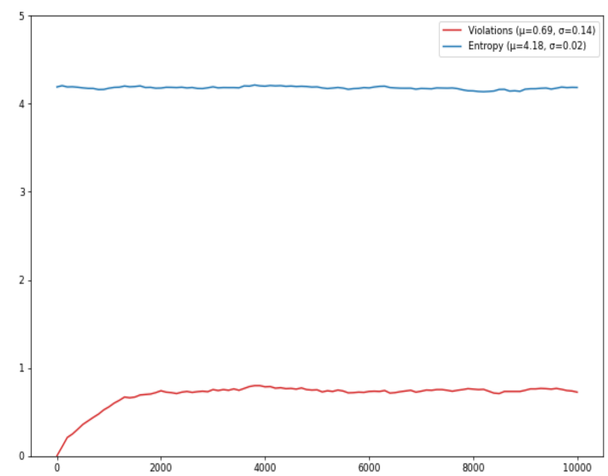
Simulation experiments

- We devised a set of computational simulations to demonstrate the effects of these alternative models of production on a Seri-like morphological system.

	NEUT SG (0)	MULT SG (1)	NEUT PL (2)	MULT PL (3)
lift (148)	**b**	**c**	**w**	**g**
bent (333)	**a**	**c**	**a**	**x**
bounce (185)	**a**	**c**	**x**	**z**
cut (111)	**a**	**c**	**w**	**g**
go to bed (37)	**a**	**c**	**x**	**y**

- A random form is deleted from a table of inflectional paradigms.
- The program tries to predict the form back, using one of three methods:
  1. **Baseline method**: copy the form used by another verb
  2. **Set-theoretic analogy**: construct and solve a set-theoretic analogy
  3. **Numeric analogy**: construct and solve an analogy using numeric features.
- This is repeated many times, cumulatively altering the system and potentially creating violations of the scale and/or affecting cell predictability.

Simulation experiments



	NEUT SG (0)	MULT SG (1)	NEUT PL (2)	MULT PL (3)
arrive	***s***	***c***	**k**	**x**
be located	**a**	**q**	***w***	***v***
cover	**g**	***s***	***k***	**r**
curved	**c**	**d**	***y***	***x***
do carefully	***s***	***c***	**k**	**x**

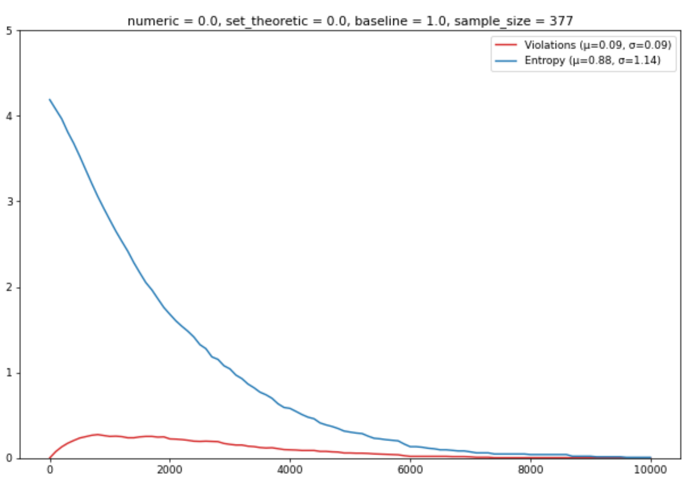
- We built in type frequency effects: the program tries to solve the problem by repeating many predictions, and choosing the majority answer. Otherwise many scale violations develop, regardless of the prediction method.

Simulation experiments

- We also built in a constraint against syncretism: if a change would make two cells the same for any given verb, it is blocked.
- This can be understood as a proxy for an anti-homophony constraint in interpretation.

	NEUT SG (0)	MULT SG (1)	NEUT PL (2)	MULT PL (3)
arrive	**s**	**s**	**s**	**s**
be located	**a**	**a**	**a**	**a**
cover	**g**	**g**	**g**	**g**
curved	**c**	**c**	**c**	**c**
do carefully	**k**	**k**	**k**	**k**

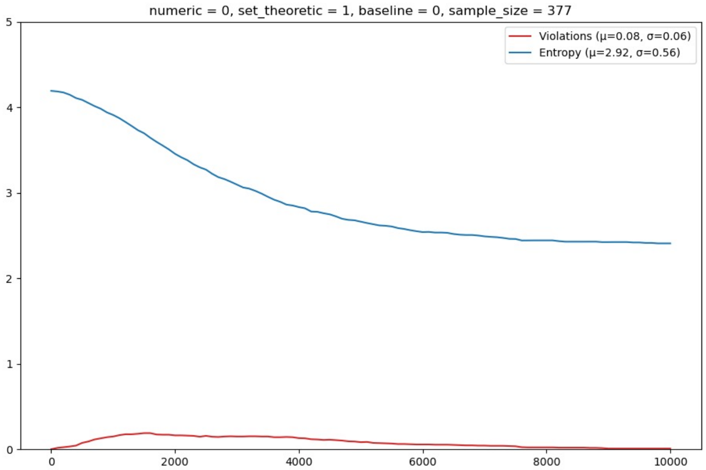
Results: baseline method



	NEUT SG (0)	MULT SG (1)	NEUT PL (2)	MULT PL (3)
arrive	**a**	**b**	**y**	**z**
be located	**a**	**b**	**y**	**z**
cover	**a**	**b**	**y**	**z**
curved	**a**	**b**	**y**	**z**
do carefully	**a**	**b**	**y**	**z**

- Scale violations initially increase, but go back to zero.
- This is achieved by generalizing a single marker for each cell, reducing entropy to zero.

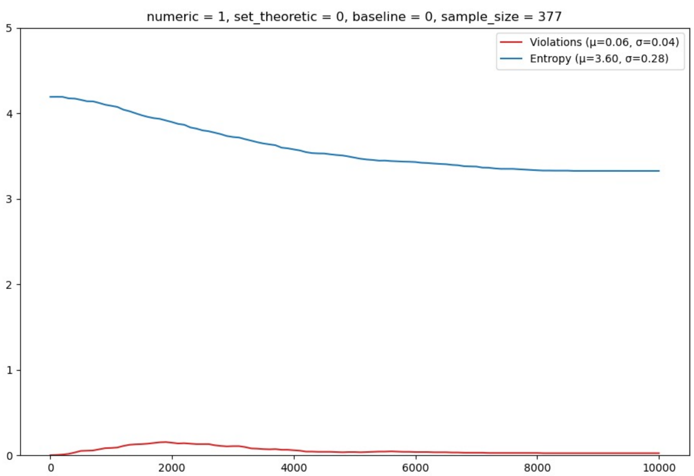
Results: set-theoretic analogy



	NEUT SG (0)	MULT SG (1)	NEUT PL (2)	MULT PL (3)
arrive	**c**	**w**	**v**	**y**
be located	**a**	**b**	**c**	**u**
cover	**b**	**c**	**u**	**v**
curved	**c**	**u**	**v**	**x**
do carefully	**a**	**b**	**c**	**u**

- Does better on scale violations than the baseline method, but not perfect.
- Entropy reduces significantly, but then remains stable.

Results: numeric analogy

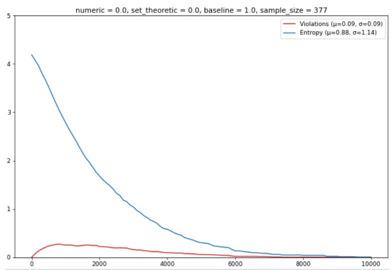


	NEUT SG (0)	MULT SG (1)	NEUT PL (2)	MULT PL (3)
arrive	**q**	**u**	**w**	**x**
be located	**b**	**c**	**d**	**e**
cover	**d**	**e**	**f**	**i**
curved	**e**	**f**	**i**	**l**
do carefully	**b**	**c**	**d**	**e**

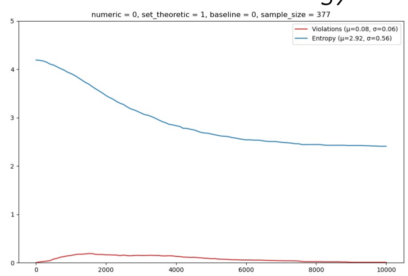
- Performs best on scale violations.
- Entropy reduces a bit, but is stable at a comparatively high level.

Conclusions

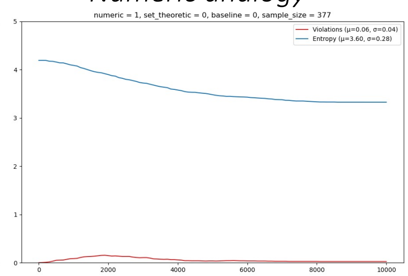
Baseline



Set-theoretic analogy



Numeric analogy



- Numeric analogy performs best at maintaining a Seri-like system: it preserves the implicational hierarchy of forms and the highest degree of allomorphy.
- Evidence for purely relational features
- The Seri system can be productively extended: speakers will produce ad-hoc forms for idiosyncratic 'extra-plural' and 'extra singular' meanings, and vice versa. This can be modelled using numeric analogical proportions.