Adaptive analogy in Word-and-Paradigm morphology

Helen Sims-Williams,[1] Jérémy Pasquereau[2] & Matthew Baerman[2]
[1] Oxford University
[2] Surrey Morphology Group, University of Surrey

Funded by the Arts & Humanities Research Council (UK) under grant AH/P002471/1 ('Seri verbs'). 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.

- Word-and-Paradigm morphology uses analogical proportions to generate forms: '**a** is to **b** as **c** is to **d**'.

a	**b**
c	**d**

- For the proportion to work, we need to be comparing like to like.
- If we want to inflect **thesis** on analogy with **crisis**:

SG	**crisis**	PL	**crises**
SG	**thesis**	PL	**thesis**

Replace the value SG with PL. ✓
Replace final **-is** with **-es**. ✓

- But what if we try to extend a morphological relationship by analogy where the features don't match?
- Imagine we want to extend the morphological relationship between plural **tables** ~ singular **table** to the verb **wobbles** in order to arrive at non-3rd person **wobble.**

PL	**tables**	SG	**table**
3	**wobbles**		

Replace the value PL with SG. ✗
Delete final **-s**.

- We can't, because there's no PL to be replaced.
- But would we ever want to construct an analogical proportion with non-matching morphosyntactic features?
  - Yes, we claim.
- How do we do this?
  - We need to replace the morphosyntactic features with something else.

Case study: Seri

- Seri is spoken on the coast of Sonora (Mexico) in two villages: El Desemboque/Haxöl Iihom and Punta Chueca/Socaaix

- It is spoken by approximately 900 speakers (Ethnologue 2007)

Seri verbs

- Number distinctions in verbs; see Marlett (1988, 2016) and Moser & Marlett (2010)
  - Subject number
    - Singular
    - Plural
  - Verbal number
    - Neutral
    - Multiple; see Cabredo Hofherr, Pasquereau, O'Meara (2018)

Seri verbs: subject number

- Singular subject

Moxima *sahmees* *hizcoi* ***h-yoohit.***
yesterday orange DEM.PL 1SG-RLS.YO.eat
'Yesterday I ate these oranges.'

Moxima *sahmees* *hizcoi* ***h-yoohitim.***
yesterday orange DEM.PL 1SG-RLS.YO.eat.MULT
'Yesterday I ate these oranges (over time).'

* *Moxima* *sahmees* *hizcoi* ***h-yoiitoj.***
yesterday orange DEM.PL 1SG-RLS.YO.eat.PL
Int. 'Yesterday I ate these oranges.'

* *Moxima* *sahmees* *hizcoi* ***h-yoiitolca.***
yesterday orange DEM.PL 1SG-RLS.YO.eat.PL.MULT
Int. 'Yesterday I ate these oranges (over time).'

Seri verbs: subject number

- Plural subject

* *Moxima* *sahmees* *hizcoi* ***ha-yoohit.***
yesterday orange DEM.PL 1PL-RLS.YO.eat
Int. 'Yesterday we ate these oranges.'

* *Moxima* *sahmees* *hizcoi* ***ha-yoohitim.***
yesterday orange DEM.PL 1PL-RLS.YO.eat.MULT
Int. 'Yesterday we ate these oranges (over time).'

Moxima *sahmees* *hizcoi* ***ha-yoiitoj.***
yesterday orange DEM.PL 1PL-RLS.YO.eat.PL
'Yesterday we ate these oranges.'

Moxima *sahmees* *hizcoi* ***ha-yoiitolca.***
yesterday orange DEM.PL 1PL-RLS.YO.eat.PL.MULT
'Yesterday we ate these oranges (over time).'

Seri verbs: event number

- Neutral form is underspecified

Juan *quih* *xiica an iqueaacalca* *coi* *hant* ***iyootox***
Juan DEF suitcases DEF.PL down 3>3.RLYO.extend
'Juan dragged the suitcases.'

Context 1: true
Distribution over times: ✓
Distribution over spaces: ?

Context 2: true
Distribution over times: ✗
Distribution over spaces: ✓

Context 3: true
Distribution over times: ✗
Distribution over spaces: ✗

Seri verbs: event number

- Neutral form is underspecified

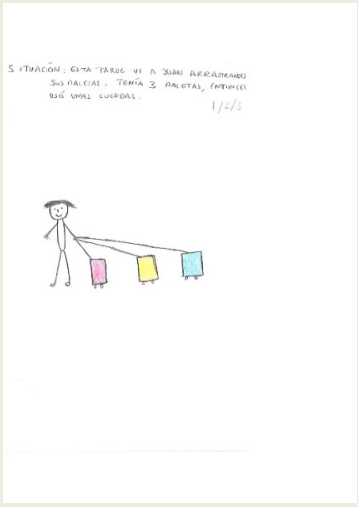
Juan *quih* *xiica an iqueaacalca*
Juan DEF suitcases
coi ***hant*** ***iyootox***
DEF.PL down 3>3.RLYO.extend
'Juan dragged the suitcases.'

Context 1: true
Distribution over times: ✓
Distribution over spaces: ?

Seri verbs: event number

- Neutral form is underspecified

Juan *quih* *xiica an iqueaacalca*
Juan DEF suitcases
coi ***hant*** ***iyootox***
DEF.PL down 3>3.RLYO.extend
'Juan dragged the suitcases.'

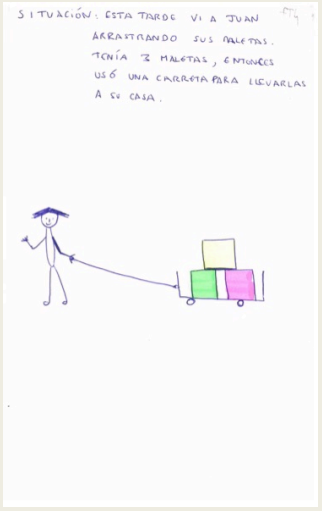


Context 2: true
Distribution over times: ✗
Distribution over spaces: ✓

Seri verbs: event number

- Neutral form is underspecified

Juan *quih* *xiica an iqueaacalca*
Juan DEF suitcases
coi ***hant*** ***iyootox***
DEF.PL down 3>3.RLYO.extend
'Juan dragged the suitcases.'



Context 3: true
Distribution over times: ✗
Distribution over spaces: ✗

Seri verbs: multiple

- Multiple form requires multiple events which are, at least, distributed over times

Juan	*quih*	*xiica an iqueaacalca*	*coi*	***hant***	***iyootoxim***
Juan	DEF	suitcases	DEF.PL	down	3>3.RLYO.extend.MULT

'Juan dragged the suitcases.'

Context 1: true

Distribution over times: ✓
Distribution over spaces: ?

Context 2: false

Distribution over times: ✗
Distribution over spaces: ✓

Context 3: false

Distribution over times: ✗
Distribution over spaces: ✗

Seri verbs

- Two cross-classifying features; see Marlett (1988, 2016) and Moser & Marlett (2010), Pasquereau & Cabredo-Hofherr (2020)

event number

subject number	neutral	multiple
singular	**iyoohit** 'one eats something'	**iyoohitim** 'one eats something over time'
plural	**iyoiitoj** 'several eat something'	**iyoiitolca** 'several eat something over time'

event number

subject number	neutral	multiple
singular	**iyopanzx** 'one runs'	**iyopanozxim** 'one runs over time'
plural	**iyopancojc** 'several run'	**iyopancoxlca** 'several run over time'

Seri verbs

- Allomorphy

event number

subject number	neutral	multiple
singular	**iyoohit** 'one eats something'	**iyoohitim** 'one eats something over time'
plural	**iyoiitoj** 'several eat something'	**iyoiitolca** 'several eat something over time'

event number

subject number	neutral	multiple
singular	**iyopanzx** 'one runs'	**iyopanozxim** 'one runs over time'
plural	**iyopancojc** 'several run'	**iyopancoxlca** 'several run over time'

No one-to-one mapping

- For every exponent$_x$, it is not possible to associate it$_x$ with one bundle of features

SG NEUTRAL	SG MULT	PL NEUTRAL	PL MULT	
-tteepx	**-tteepx-tim**	**-tteepzaj-c**	**-tteepzal-ca**	'sit on'
-taxnij	**-taxanl**	**-taxnal-ca**	**-taxnal-coj**	'scold'
-tanamj	**-tanaml-c**	**anaml-coj**	**anaml-cam**	'hurry'
-tazaain-im	**-tazaiin-im**	**azaail-cam**	**azaiil-cam**	'anchor'
-tahipxa	**-tapxaal-ca**	**-tahipxal-ca**	**-tahipxal-ca**	'roll up'
-tpoc	**-tpoct-im**	**-tpocl-im**	**-tpocal-am**	'fall'
-tpazj-c	**-tpaxlax**	**-tpazlax**	**-tpazlax-lca**	'be scattered'
-tineezil-ca	**-tineezil-im**	**-tineezil-coj**	**-tineezil-am**	'be raspy'

-tim

No one-to-one mapping

- For every exponent$_x$, it is not possible to associate it$_x$ with one bundle of features

SG NEUTRAL	SG MULT	PL NEUTRAL	PL MULT	
-tteepx	**-tteepx-tim**	**-tteepzaj-c**	**-tteepzal-ca**	'sit on'
-taxnij	**-taxanl**	**-taxnal-ca**	**-taxnal-coj**	'scold'
-tanamj	**-tanaml-c**	**anaml-coj**	**anaml-cam**	'hurry'
-tazaain-im	**-tazaiin-im**	**azaail-cam**	**azaiil-cam**	'anchor'
-tahipxa	**-tapxaal-ca**	**-tahipxal-ca**	**-tahipxal-ca**	'roll up'
-tpoc	**-tpoct-im**	**-tpocl-im**	**-tpocal-am**	'fall'
-tpazj-c	**-tpaxlax**	**-tpazlax**	**-tpazlax-lca**	'be scattered'
-tineezil-ca	**-tineezil-im**	**-tineezil-coj**	**-tineezil-am**	'be raspy'

-c

No one-to-one mapping

- For every exponent$_x$, it is not possible to associate it$_x$ with one bundle of features

SG NEUTRAL	SG MULT	PL NEUTRAL	PL MULT	
-tteepx	**-tteepx-tim**	**-tteepzaj-c**	**-tteepzal-ca**	'sit on'
-taxnij	**-taxanl**	**-taxnal-ca**	**-taxnal-coj**	'scold'
-tanamj	**-tanaml-c**	**anaml-coj**	**anaml-cam**	'hurry'
-tazaain-im	**-tazaiin-im**	**azaail-cam**	**azaiil-cam**	'anchor'
-tahipxa	**-tapxaal-ca**	**-tahipxal-ca**	**-tahipxal-ca**	'roll up'
-tpoc	**-tpoct-im**	**-tpocl-im**	**-tpocal-am**	'fall'
-tpazj-c	**-tpaxlax**	**-tpazlax**	**-tpazlax-lca**	'be scattered'
-tineezil-ca	**-tineezil-im**	**-tineezil-coj**	**-tineezil-am**	'be raspy'

-ca

No one-to-one mapping

- For every exponent$_x$, it is not possible to associate it$_x$ with one bundle of features

SG NEUTRAL	SG MULT	PL NEUTRAL	PL MULT	
-tteepx	**-tteepx-tim**	**-tteepzaj-c**	**-tteepzal-ca**	'sit on'
-taxnij	**-taxanl**	**-taxnal-ca**	**-taxnal-coj**	'scold'
-tanamj	**-tanaml-c**	**anaml-coj**	**anaml-cam**	'hurry'
-tazaain-im	**-tazaiin-im**	**azaail-cam**	**azaiil-cam**	'anchor'
-tahipxa	**-tapxaal-ca**	**-tahipxal-ca**	**-tahipxal-ca**	'roll up'
-tpoc	**-tpoct-im**	**-tpocl-im**	**-tpocal-am**	'fall'
-tpazj-c	**-tpaxlax**	**-tpazlax**	**-tpazlax-lca**	'be scattered'
-tineezil-ca	**-tineezil-im**	**-tineezil-coj**	**-tineezil-am**	'be raspy'

-coj

No one-to-one mapping

- For every exponent$_x$, it is not possible to associate it$_x$ with one bundle of features

SG NEUTRAL	SG MULT	PL NEUTRAL	PL MULT	
-tteepx	**-tteepx-tim**	**-tteepzaj-c**	**-tteepzal-ca**	'sit on'
-taxnij	**-taxanl**	**-taxnal-ca**	**-taxnal-coj**	'scold'
-tanamj	**-tanaml-c**	**anaml-coj**	**anaml-cam**	'hurry'
-tazaain-im	**-tazaiin-im**	**azaail-cam**	**azaiil-cam**	'anchor'
-tahipxa	**-tapxaal-ca**	**-tahipxal-ca**	**-tahipxal-ca**	'roll up'
-tpoc	**-tpoct-im**	**-tpocl-im**	**-tpocal-am**	'fall'
-tpazj-c	**-tpaxlax**	**-tpazlax**	**-tpazlax-lca**	'be scattered'
-tineezil-ca	**-tineezil-im**	**-tineezil-coj**	**-tineezil-am**	'be raspy'

-tim **-c** **-ca** **-coj**

Incrementality of the exponents

- The system is not completely random however, because both the paradigm cells and the exponents *mostly* follow an implicational scale :

singular ⟶ plural

SG NEUTRAL	SG MULT	PL NEUTRAL	PL MULT

-(t)im	**-c**	**-ca**	**-lca**	**-coj**	**-(c)am**

singular ⟶ plural

Incrementality of the exponents

- The system is not completely random however, because both the paradigm cells and the exponents *mostly* follow an implicational scale :

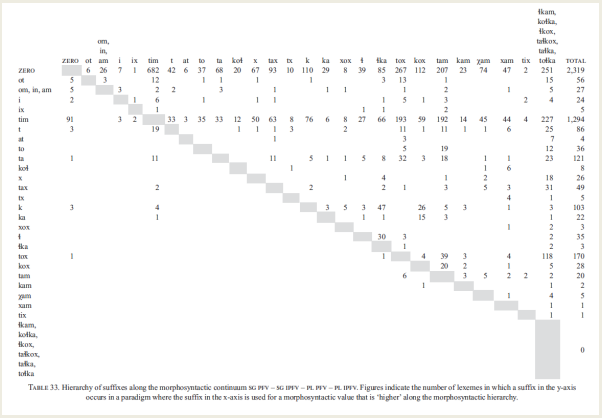


TABLE 33. Hierarchy of suffixes along the morphosyntactic continuum SG IPFV – SG IPFV – PL PFV – PL IPFV. Figures indicate the number of lexemes in which a suffix in the y-axis occurs in a paradigm where the suffix in the x-axis is used for a morphosyntactic value that is 'higher' along the morphosyntactic hierarchy.

Incrementality of the exponents

- The system is not completely random however, because both the paradigm cells and the exponents *mostly* follow an implicational scale :

singular ⟶ plural

SG NEUTRAL	SG MULT	PL NEUTRAL	PL MULT

-(t)im	**-c**	**-ca**	**-lca**	**-coj**	**-(c)am**

singular ⟶ plural

SG NEUTRAL	SG MULT	PL NEUTRAL	PL MULT	
-tteepx	**-tteepx-tim**	**-tteepzaj-c**	**-tteepzal-ca**	'sit on'

Incrementality of the exponents

- The system is not completely random however, because both the paradigm cells and the exponents *mostly* follow an implicational scale :

singular ⟶ plural

SG NEUTRAL	SG MULT	PL NEUTRAL	PL MULT

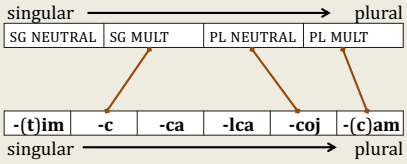
-(t)im	**-c**	**-ca**	**-lca**	**-coj**	**-(c)am**

singular ⟶ plural

SG NEUTRAL	SG MULT	PL NEUTRAL	PL MULT	
-taxnij	**-taxanl**	**-taxnal-ca**	**-taxnal-coj**	'scold'

Incrementality of the exponents

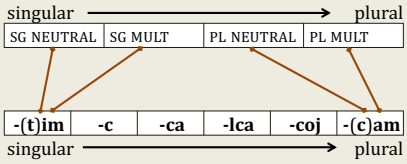
- The system is not completely random however, because both the paradigm cells and the exponents *mostly* follow an implicational scale :



SG NEUTRAL	SG MULT	PL NEUTRAL	PL MULT	
-tanamj	**-tanaml-c**	**anaml-coj**	**anaml-cam**	'hurry'

Incrementality of the exponents

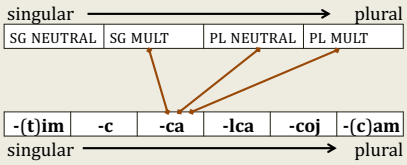
- The system is not completely random however, because both the paradigm cells and the exponents *mostly* follow an implicational scale :



SG NEUTRAL	SG MULT	PL NEUTRAL	PL MULT	
-tazaain-im	**-tazaiin-im**	**azaail-cam**	**azaiil-cam**	'anchor'

Incrementality of the exponents

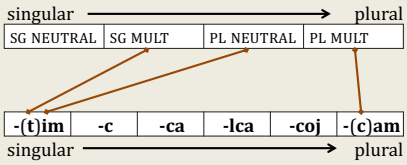
- The system is not completely random however, because both the paradigm cells and the exponents *mostly* follow an implicational scale :



SG NEUTRAL	SG MULT	PL NEUTRAL	PL MULT	
-tahipxa	**-tapxaal-ca**	**-tahipxal-ca**	**-tahipxal-ca**	'roll up'

Incrementality of the exponents

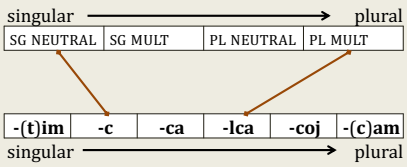
- The system is not completely random however, because both the paradigm cells and the exponents *mostly* follow an implicational scale :



SG NEUTRAL	SG MULT	PL NEUTRAL	PL MULT	
-tpoc	**-tpoct-im**	**-tpocl-im**	**-tpocal-am**	'fall'

Incrementality of the exponents

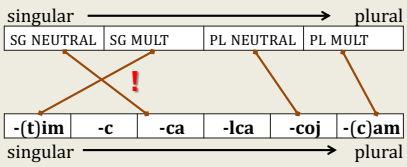
- The system is not completely random however, because both the paradigm cells and the exponents *mostly* follow an implicational scale :



SG NEUTRAL	SG MULT	PL NEUTRAL	PL MULT	
-tpazj-c	**-tpaxlax**	**-tpazlax**	**-tpazlax-lca**	'be scattered'

Incrementality of the exponents

- The system is not completely random however, because both the paradigm cells and the exponents *mostly* follow an implicational scale :



SG NEUTRAL	SG MULT	PL NEUTRAL	PL MULT	
-tineezil-ca	**-tineezil-im**	**-tineezil-coj**	**-tineezil-am**	'be raspy'

-(t)im is an outlier

Where does the hierarchy come from?

- Internal evidence suggests concatenation of plural marking at some point in history.

SG NEUTRAL	SG MULT	PL NEUTRAL	PL MULT	
-iiquet	**-iicto**	**-iictoj**	**-iictolca**	'be pregnant'

- Compare:

SG NEUTRAL	PL NEUTRAL	
-eeme	**-eeme-t**	'use up'
-oonl	**-oonl-o**	'stir'
-axaa	**-axaa-j**	'gather roots'
-ihapoj	**-ihapol-ca**	'have digging stick'

Note: **j → l /__ c**

- Thus the "more plural" a form is, the more plural marking it has.
- No reason to think this principle is still productive, but its residue colors the entire system.

Computational model

- Many-to many form-property mapping precludes storage as individual morphemes.

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

- We assume inflected forms can be stored whole in memory – but we still need a mechanism for production other than retrieval from memory:

  - Memory isn't perfect
  - The system can be creatively extended (see appendix)

- Can analogy account for this, while maintaining the Baerman scale?

The analogical method

- Uses an algorithm by Lepage (1998) to solve analogical proportions by matching shared sequences of symbols between the items in a proportion.

	crisis	**crises**
	thesis	**x = ?**

The analogical method

- Uses an algorithm by Lepage (1998) to solve analogical proportions by matching shared sequences of symbols between the items in a proportion.

	is	**es**
cris	**crisis**	**crises**
thes	**thesis**	**x = ?**

The analogical method

- A function converts morphosyntactic property sets to points on a numeric scale.

Plurality = 0
If subject number = plural: add 2 to plurality
If event number = multiple: add 1 to plurality
Yield plurality

Cell:	SG NEUT	SG MULT	PL NEUT	PL MULT
Plurality:	**0**	**1**	**2**	**3**

- Instead of working with the morphosyntactic properties directly, analogy works with these numeric values.

The analogical method

- This allows us to overcome the fact that the morphosyntactic feature values do not necessarily match, and look instead at their relative positions in the paradigm.

Replace the value SG MULT with PL NEUT ✗
Replace final **-c** with **-coj**

SG MULT **itanamlc**	PL NEUT **itanamlcoj**
PL NEUT **tmaasilc**	

The analogical method

- This allows us to overcome the fact that the morphosyntactic feature values do not necessarily match, and look instead at their relative positions in the paradigm.

Add +1 to the numerical value ✓
Replace final **-c** with **-coj** ✓

1 **itanamlc**	2 **itanamlcoj**
2 **tmaasilc**	3 **tmaasilcoj**

The analogical method

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

0 (SG NEUT)	1 (SG MULT)	2 (PL NEUT)	3 (PL MULT)	
itanamj	**itanamlc**	**itanamlcoj**	**itanamlcam**	'hurry'
tmaasij	**tmaasilim**	**tmaasilc**		'roll'
etc…				

- The program tries to predict the form by randomly sampling form-value pairings from the lexicon and putting them into an analogical proportion, until it finds one that leads to a solution:

itanamlc (1) is to **itanamlcoj** (2) as **tmaasilc** (2) is to **tmaasilcoj** (3)

Majority-rules analogical method

- As above, but the process is repeated 60 times and the result is stored each time.
- At the end, the most common result is passed forward.

Baseline method

- Has information about the relative frequency of each suffix in each paradigm cell.
- Predicts the suffix of a form with probability commensurate to type frequency.

Results (1000 Trials)

	Form predicted correctly (%)	Suffix predicted correctly (%)	Conforms to scale (%)	Scale not applicable (%)
0. Baseline method	N/A	38.4%	91.9%	0.0%
1. Analogical method	5.1%	19.6%	92.9%	37.7%
2. Majority-rules method	15.7%	28.5%	99.2%	17.2%

- Overall, the baseline method performs best for predicting the suffix correctly (but it doesn't predict whole forms).
- But the analogical methods perform better for avoiding scale violations, especially the majority-rules method.

Results (singular cells)

	Form predicted correctly (%)	Suffix predicted correctly (%)	Conforms to scale (%)	Scale not applicable (%)
0. Baseline method	N/A	68.3%	98.1%	0.0%
1. Analogical method	5.3%	28.6%	94.8%	35.2%
2. Majority-rules method	9.2%	24.4%	99.5%	17.1%

- Baseline method performs best for predicting singular cells.

Results (plural cells)

	Form predicted correctly (%)	Suffix predicted correctly (%)	Conforms to scale (%)	Scale not applicable (%)
0. Baseline method	N/A	13.3%	86.5%	0.0%
1. Analogical method	5.0%	11.5%	91.3%	42.5%
2. Majority-rules method	22.9%	28.4%	99.7%	14.0%

- Baseline method performs poorly for predicting plural cells.

Why?

	Singular subject	Plural subject
Neutral	Entropy: 1.05 bits Most common suffix: -# (85% of verbs)	Entropy: 3.76 bits Most common suffix: **-toj** (25% of verbs)
Multiple	Entropy: 1.36 bits Most common suffix: **-tim** (79% of verbs)	Entropy: 3.39 bits Most common suffix: **-tolca** (29% of verbs)

- On their own, the suffixes of the singular cells are much more predictable than those of the plural cells.

Why?

Cell 1	Cell 2	H(Cell 2|Cell 1)	H(Cell 2)	H(Cell 2)-H(Cell 2|Cell 1)
PL NEUT	PL MULT	1.534	3.393	1.859
PL MULT	PL NEUT	1.897	3.756	1.859
SG MULT	PL NEUT	3.269	3.756	0.486
PL NEUT	SG MULT	0.869	1.355	0.486
SG NEUT	PL NEUT	3.411	3.756	0.345
PL NEUT	SG NEUT	0.702	1.047	0.345
SG MULT	PL MULT	3.051	3.393	0.341
PL MULT	SG MULT	1.014	1.355	0.341
SG NEUT	PL MULT	3.093	3.393	0.300
PL MULT	SG NEUT	0.747	1.047	0.300
SG NEUT	SG MULT	1.167	1.355	0.188
SG MULT	SG NEUT	0.858	1.047	0.188

- Plural cells are mutually informative, while singular cells are not.

Conclusion

- Seri data show that analogies may obtain between forms realizing (synchronically) incongruent features.
- We represent this by replacing morphosyntactic feature values with abstract paradigmatic relationships: the formal relationships remain constant, even where the morphosyntax varies.
- In Seri this takes the form of numerical scale.

SG NEUT	SG MULT	PL NEUT	PL MULT	*morphosyntax*
0	1	2	3	*morphological paradigm structure*

- We understand this scale to be part of morphology, describing a linear paradigmatic arrangement. Obviously though it has SOME relationship to incrementally added plurality. (Whether synchronic or diachronic is another question.)

- We can then extend this approach to other contexts where a consistent paradigmatic relationship across constexts in spite of differences in the morphosyntax.
- Polarity (Meinhof 1912, Hetzron 1967): flipping of values/forms.

Hebrew

FEM	MASC	
tov-a	**tov-Ø**	adective: 'good'
šaloš-Ø	**šloš-a**	numeral: 'three'

FEM	MASC	*morphosyntax*
A	¬A	*morphological paradigm structure*