

Simulating paradigm evolution: analogical change and morphomic patterns

Sacha Beniamine, Louise Esher, Erich Round



January 23, 2020

How do inflectional paradigms change ?

Latin, present tense

	1SG	2SG	3SG	1PL	2PL	3PL
CANTARE	kanto:	kanta:s	kantat	kanta:mus	kanta:tis	kantant
DICERE	di:ko:	di:kis	di:kit	di:kimus	di:kitis	di:kunt
FACERE	fakio:	fakis	fakit	fakimus	fakitis	fakiunt
TENERE	teneo:	tene:s	tenet	tene:mus	tene:tis	tenent
UENIRE	wenio:	weni:s	wenit	weni:mus	weni:tis	weniunt

How do inflectional paradigms change ?

Modern standard Romanian, present tense

	1SG	2SG	3SG	1PL	2PL	3PL
CANTARE	'kint	'kints ⁱ	'kintə	kin'təm	kin'tats ⁱ	'kintə
DICERE	'zik	'zitʃ	'zitʃe	'zitʃem	'zitʃets ⁱ	'zik
FACERE	'fak	'fatʃ	'fatʃe	'fatʃem	'fatʃets ⁱ	'fak
TENERE	'tsin	'tsii	'tsine	'tsinem	'tsinets ⁱ	'tsin
UENIRE	'vin	'vii	'vine	ve'nim	ve'nits ⁱ	'vin

Modern standard Spanish, present tense

	1SG	2SG	3SG	1PL	2PL	3PL
CANTARE	'kanto	'kantas	'kanta	kan'tamos	kan'tajs	'kantan
DICERE	'diyo	'diθes	'diθe	de'θimos	de'θis	'diθen
FACERE	'ayo	'aθes	'aθe	a'θemos	a'θejs	'aθen
TENERE	'tenɣo	'tjenes	'tjene	te'nemos	te'nejs	'tjenen
UENIRE	'benɣo	'bjenes	'bjene	be'nimos	be'nis	'bjenen

Beniamine, Maiden, and Round (2020) based on the Oxford Romance Database (2010))

Morphemes

	1SG	2SG	3SG	1PL	2PL	3PL
CANTARE	'kanto	'kantas	'kanta	kan'tamos	kan'tajs	'kantan
DICERE	'diyo	'diθes	'diθe	de'θimos	de'θis	'diθen
FACERE	'ayo	'aθes	'aθe	a'θemos	a'θejs	'aθen
TENERE	'tengo	'tjenes	'tjene	te'nemos	te'nejs	'tjenen
UENIRE	'bengo	'bjenes	'bjene	be'nimos	be'nis	'bjenen

Modern standard Spanish, present tense

■ Morphemes (Aronoff, 1994)

- Inflection classes: declensions and conjugation groups.

Morphemes

	1SG	2SG	3SG	1PL	2PL	3PL
CANTARE	'kanto	'kantas	'kanta	kan'tamos	kan'tajs	'kantan
DICERE	'diyo	'diθes	'diθe	de'θimos	de'θis	'diθen
FACERE	'ayo	'aθes	'aθe	a'θemos	a'θejs	'aθen
TENERE	'tengo	'tjenes	'tjene	te'nemos	te'nejs	'tjenen
UENIRE	'benjo	'bjenes	'bjene	be'nimos	be'nis	'bjenen

Modern standard Spanish, present tense

■ Morphemes (Aronoff, 1994)

- Inflection classes: declensions and conjugation groups.
- Metamorphemes (Round, 2015): such as stem spaces

Morphemes

	1SG	2SG	3SG	1PL	2PL	3PL
CANTARE	'kanto	'kantas	'kanta	kan'tamos	kan'tajs	'kantan
DICERE	'diyo	'diθes	'diθe	de'θimos	de'θis	'diθen
FACERE	'ayo	'aθes	'aθe	a'θemos	a'θejs	'aθen
TENERE	'tengo	'tjenes	'tjene	te'nemos	te'nejs	'tjenen
UENIRE	'benjo	'bjenes	'bjene	be'nimos	be'nis	'bjenen

Modern standard Spanish, present tense

■ Morphemes (Aronoff, 1994)

- Inflection classes: declensions and conjugation groups.
- Metamorphemes (Round, 2015): such as stem spaces

■ They persist and spread (Maiden, 2018):

- Why ? Recurrent patterns are predictable (Carstairs-McCarthy, 2010; Esher, 2015b; Blevins, 2016)

Morphemes

	1SG	2SG	3SG	1PL	2PL	3PL
CANTARE	'kanto	'kantas	'kanta	kan'tamos	kan'tajs	'kantan
DICERE	'diyo	'diθes	'diθe	de'θimos	de'θis	'diθen
FACERE	'ayo	'aθes	'aθe	a'θemos	a'θejs	'aθen
TENERE	'tengo	'tjenes	'tjene	te'nemos	te'nejs	'tjenen
UENIRE	'benngo	'bjenes	'bjene	be'nimos	be'nis	'bjenen

Modern standard Spanish, present tense

■ Morphemes (Aronoff, 1994)

- Inflection classes: declensions and conjugation groups.
- Metamorphemes (Round, 2015): such as stem spaces

■ They persist and spread (Maiden, 2018):

- Why ? Recurrent patterns are predictable (Carstairs-McCarthy, 2010; Esher, 2015b; Blevins, 2016)
- How ? By analogy

Mechanisms of paradigm change

DICERE

	1SG	2SG	3SG	1PL	2PL	3PL
Latin						
PRS.IND	di:ko:	di:kis	di:kit	di:kimus	di:kitis	di:kunt
PRS.SBJV	di:kam	di:ka:s	di:kat	di:ka:mus	di:ka:tis	di:kant
Spanish						
PRS.IND	'diyo	'diθes	'diθe	de'θimos	de'θis	'diθen
PRS.SBJV	'diya	'diyas	'diya	di'ɣamos	di'ɣajs	'diyan

- **Sound changes** affect phonemes in specific contexts

Mechanisms of paradigm change

DICERE

	1SG	2SG	3SG	1PL	2PL	3PL
Latin						
PRS.IND	di:ko:	di:kis	di:kit	di:kimus	di:kitis	*di:kent
PRS.SBJV	di:kam	di:ka:s	di:kat	di:ka:mus	di:ka:tis	di:kant
Spanish						
PRS.IND	'diyo	'diθes	'diθe	de'θimos	de'θis	'diθen
PRS.SBJV	'diya	'diyas	'diya	di'ɣamos	di'ɣajs	'diyan

- **Sound changes** affect phonemes in specific contexts
- **Analogical changes** remodels forms by copying alternation patterns.

Outline for this talk

What are the analogical processes which lead to the appearance, persistence and spread of morphemes ?

1. Two interpretations of analogical change
2. Two very simple simulations of analogical change
3. Simulating the evolution of L/U patterns
4. Conclusion

Predicting unknown cells

Paradigm Cell Filling Problem

What licenses reliable inferences about surface word forms ?
(Ackerman and Malouf, 2013)

- Paradigm forms follow a zipfian distribution (Bonami and Beniamine, 2016, Blevins and al. 2017)
- Speakers fill unknown forms by relying on strong **analogical** relations
- measured by Ackerman and Malouf (2013, 2015) and Bonami and Beniamine (2016) using entropy.

Paradigm evolution driven by the PCFP

English

	INF	PRS.3SG	PST	PST.PTCP	PRS.PTCP
LIVE	liv	livz	livd	livd	livɪŋ
DIVE	daɪv	daɪvz	daɪvd	daɪvd	daɪvɪŋ
DRIVE	draɪv	draɪvz	drəʊv	drɪvɪŋ	draɪvɪŋ
RIDE	raɪd	raɪdz	rəʊd	rɪdɪŋ	raɪdɪŋ

- Analogical change happening in the context of the PCFP

Paradigm evolution driven by the PCFP

English

	INF	PRS.3SG	PST	PST.PTCP	PRS.PTCP
LIVE	liv	livz	livd	livd	livɪŋ
DIVE	daɪv	daɪvz	dəʊv	daɪvd	daɪvɪŋ
DRIVE	draɪv	draɪvz	drəʊv	drɪvɪŋ	draɪvɪŋ
RIDE	raɪd	raɪdz	rəʊd	rɪdɪŋ	raɪdɪŋ

- Analogical change happening in the context of the PCFP

Paradigm evolution driven by the PCFP

English

	INF	PRS.3SG	PST	PST.PTCP	PRS.PTCP
LIVE	lɪv	lɪvz	lɪvd	lɪvd	lɪvɪŋ
DIVE	dəɪv	dəɪvz	dəʊv	dəɪvd	dəɪvɪŋ
DRIVE	draɪv	draɪvz	drəʊv	drɪvɪŋ	draɪvɪŋ
RIDE	raɪd	raɪdz	rəʊd	rɪdɪŋ	raɪdɪŋ

- Analogical change happening in the context of the PCFP
- Ackerman and Malouf (2015) ran simulations where, starting from random paradigms, a model makes this type of analogical change until the system is stable.

Metamorphomes in Romance verbs

	1SG	2SG	3SG	1PL	2PL	3PL
TENERE						
PRS.IND	teneo:	tene:s	tenet	tene:mus	tene:tis	tenent
PRS.SBJV	teneam	tenea:s	teneat	tenea:mus	tenea:tis	teneant
DICERE						
PRS.IND	di:ko:	di:kis	di:kit	di:kimus	di:kitis	di:kunt
PRS.SBJV	di:kam	di:ka:s	di:kat	di:ka:mus	di:ka:tis	di:kant

- L and U patterns formed by sound changes (Maiden, 2018)
 - Yod Effect
 - Palatalization and Affrication of velars

Metamorphomes evolution

	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	teneo:	tene:s	tenet	tene:mus	tene:tis	tenent
PRS.SBJV	teneam	tenea:s	teneat	tenea:mus	tenea:tis	teneant

Latin, TENERE

PRS.IND	'tengo	'tjenes	'tjene	te'nemos	te'nejs	'tjenen
PRS.SBJV	'tenga	'tengas	'tenga	ten'gamos	ten'gajs	'tengan

Spanish, TENERE

- Maiden (2018): These are often preserved

Metamorphomes evolution

	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	di:ko:	di:kis	di:kit	di:kimus	di:kitis	di:kunt
PRS.SBJV	di:kam	di:ka:s	di:kat	di:ka:mus	di:ka:tis	di:kant

Latin, DICERE

PRS.IND	'diyo	'diθes	'diθe	de'θimos	de'θis	'diθen
PRS.SBJV	'diya	'diyas	'diya	di'yamos	di'yajs	'diyan

Spanish, DICERE

- Maiden (2018): These are often preserved and extended

Metamorphomes evolution

	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	possum	potes	potest	possumus	potestis	possunt
PRS.SBJV	possim	possis	possit	possimus	possitis	possint

Latin, POSSE

PRS.IND	'posu	'pɔdɨ	'pɔdɨ	pu'demuɟ	pu'deɟɨ	'pɔdẽĩ
PRS.SBJV	'pose	'poseɟ	'pose	pu'semuɟ	pu'saɟɨ	'posẽũ

Portuguese, POSSE

- Maiden (2018): These are often preserved and extended, even to different types of alternations

"We can be sure that speakers have made such generalizations because the L/U pattern turns out to be a historical leitmotif of the romance verb."

Maiden (2018)

Modelling analogical change

What kind of analogical change replicate morphomic patterns ?
Can identity analogy improve predictability ?

	A	B	C
I1	Light Green	Light Green	Light Green
I2	Purple	Purple	Blue
I3	Light Green	?	Red
I4	Light Green	Blue	Red
I5	Light Green	Blue	Blue

- Starting with random paradigms,
 1. Hold out one value,
 2. Predict it using an analogical model.
 3. Replace the previous value
- This was 1 cycle. Repeat until stable

Random paradigms



- 100 lexemes (rows)
- 8 paradigm cells (columns)
- A total pool of 6 possible allomorphs (numbers)
- Each cell has only 3 allomorphs

Substance based analogy

	A	B	C
l1	light green	light green	light green
l2	purple	purple	blue
l3	light green	?	red
l4	light green	blue	red
l5	light green	blue	blue

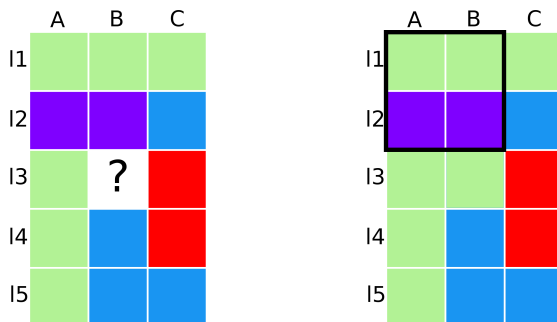
	A	B	C
l1	light green	light green	light green
l2	purple	purple	blue
l3	light green	blue	red
l4	light green	blue	red
l5	light green	blue	blue

$$\operatorname{argmax}_x P(B = x | A = y)$$

- Pick a lexeme l and two cells A, B . We predict l^B knowing A
- Select the lexemes with the same value as l in cell A
- Predict the most likely value in B from these lexemes.

inspired by Ackerman and Malouf (2015)

Identity based analogy

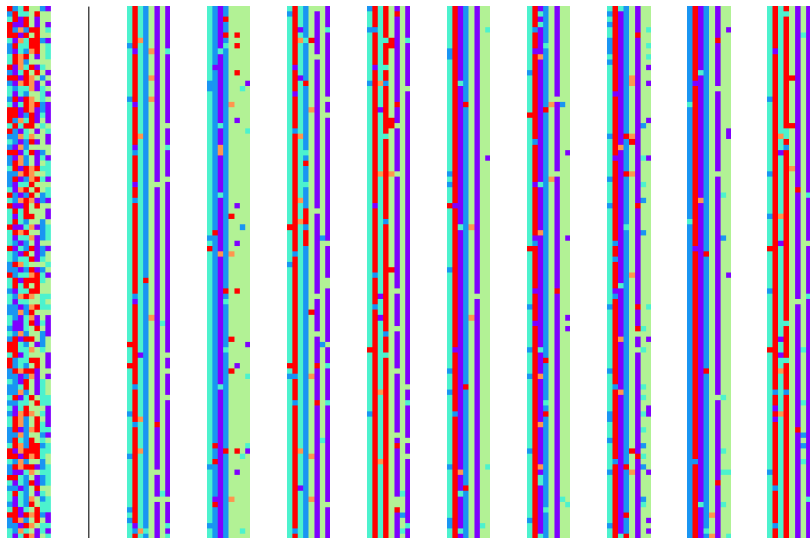


$$\operatorname{argmax}_X P(X == B)$$

- Pick a lexeme l and a cell B . We predict l^B .
- Select the cell X with the highest probability to be identical to B
- Copy the value of l from X to B

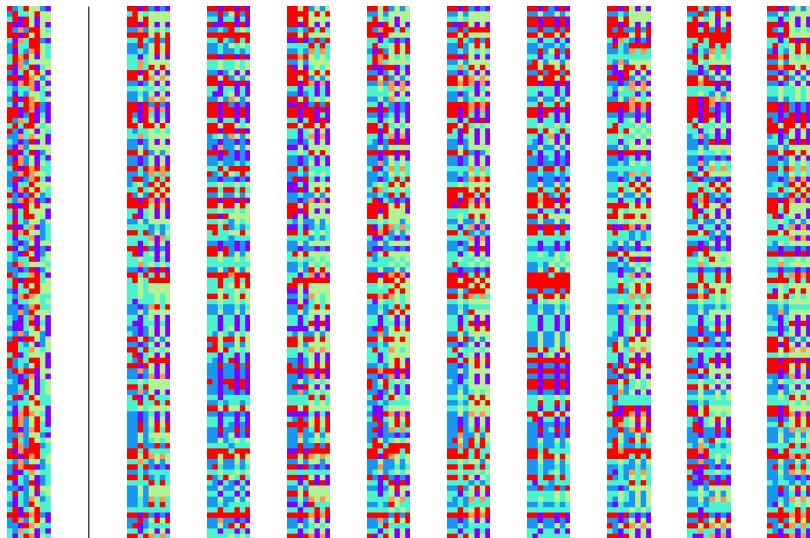
inspired by Esher (2015b,a, 2016, 2017)

Results: Substance analogy



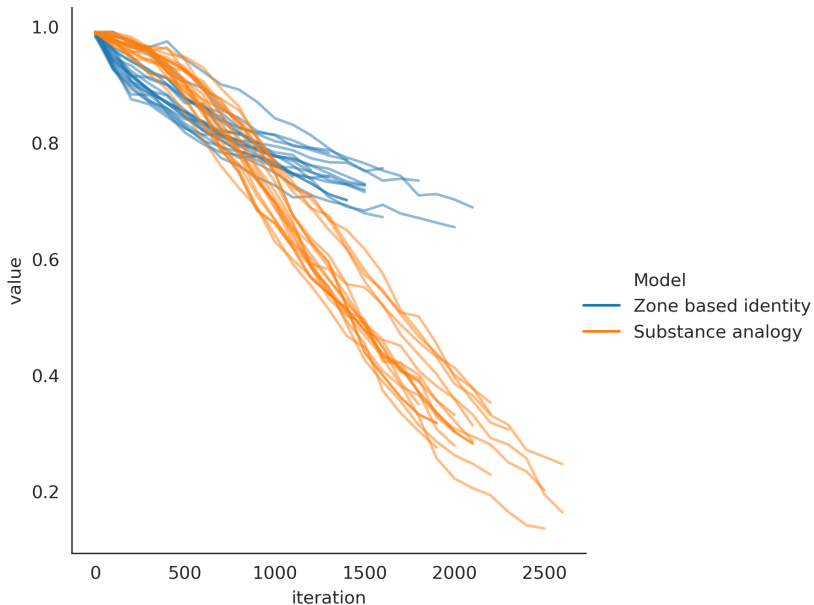
We run each model 20 times

Results: Identity analogy

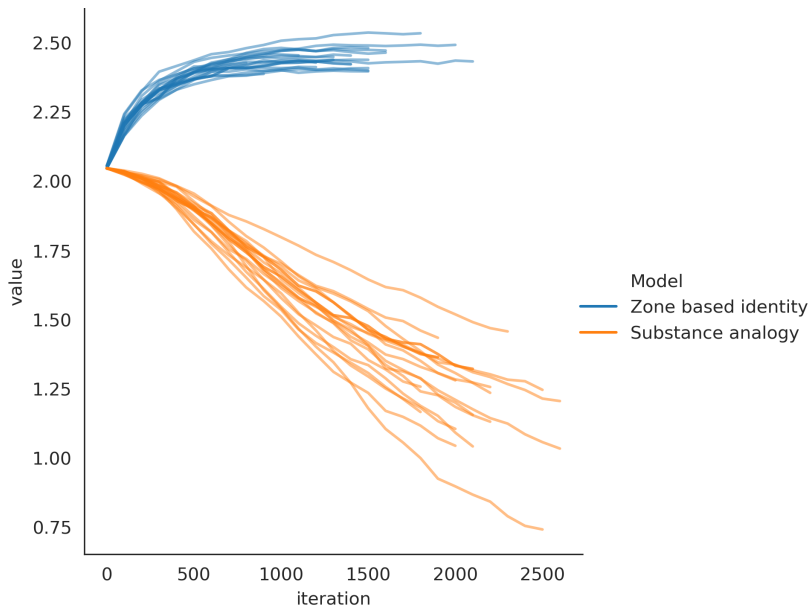


We run each model 20 times

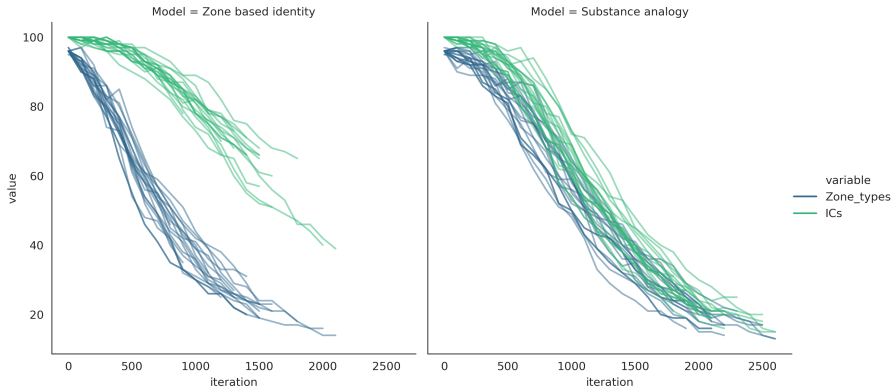
Paradigm entropy



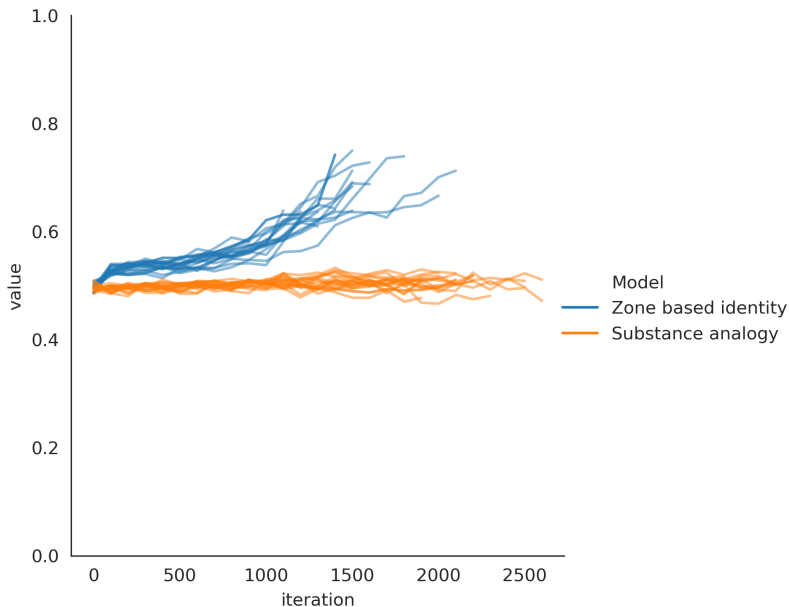
Cell entropy



Morphome evolution



Metamorphomes: hopkins statistic



Conclusion on random models

- Identity analogy is necessary to obtain metamorphomes
 - Conformity bias is crucial ! (contrast to Esher, 2017)
 - We should be careful of assuming substance analogy when measuring the PCFP !
- The metaphors we entertain when dealing with artificial data have an impact on our measurements

Can these spread the L/U patterns ?

	1SG	2SG	3SG	1PL	2PL	3PL
TENERE						
PRS.IND	teneo:	tene:s	tenet	tene:mus	tene:tis	tenent
PRS.SBJV	teneam	tenea:s	teneat	tenea:mus	tenea:tis	teneant
DICERE						
PRS.IND	di:ko:	di:kis	di:kit	di:kimus	di:kitis	di:kunt
PRS.SBJV	di:kam	di:ka:s	di:kat	di:ka:mus	di:ka:tis	di:kant
AMARE						
PRS.IND.	amo:	ama:s	amat	ama:mus	ama:tis	amant
PRS.SBJ.	amem	ame:s	amet	ame:mus	ame:tis	ament

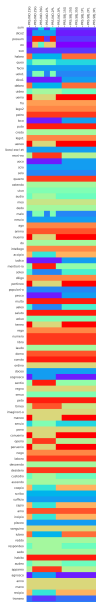
- From LatinFlexi (Pellegrini and Passarotti, 2018), keep:
 - 100 lexemes and 12 cells

Can these spread the L/U patterns ?

	1SG	2SG	3SG	1PL	2PL	3PL
TENERE						
PRS.IND	n ^j	n	n	n	n	n
PRS.SBJV	n ^j	n ^j	n ^j	n ^j	n ^j	n ^j
DICERE						
PRS.IND	k	k ^j	k ^j	k ^j	k ^j	k
PRS.SBJV	k	k	k	k	k	k
AMARE						
PRS.IND.	m	m	m	m	m	m
PRS.SBJ.	m	m	m	m	m	m

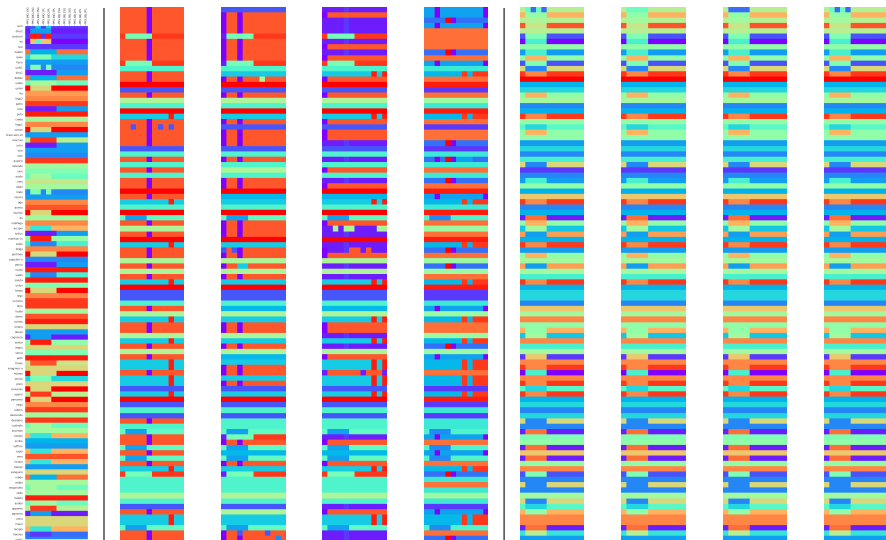
- From LatinFlexi (Pellegrini and Passarotti, 2018), keep:
 - 100 lexemes and 12 cells
 - the last stem C, marking YE / PAV environments by [ʲ]

Can these spread the L/U patterns ?



- From LatinFlexi (Pellegrini and Passarotti, 2018), keep:
 - 100 lexemes and 12 cells
 - the last stem C, marking YE / PAV environments by [ʲ]
- Code this on integers (colors)

Results



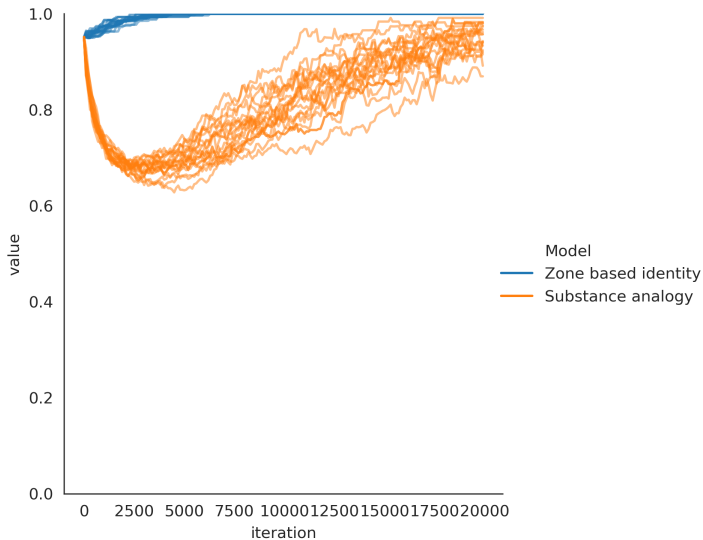
initial

substance analogy

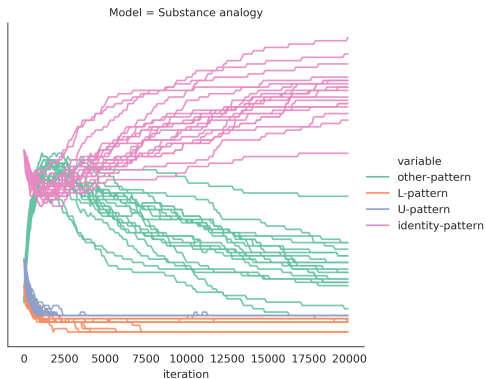
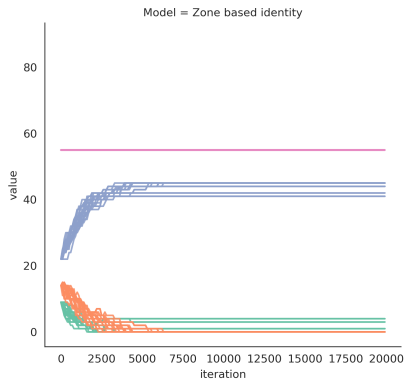
identity analogy

Metamorphomes: hopkins statistic

U/L dataset



Evolution of L/U patterns



What next ?

This is a work in progress. Possible improvements:

- A mixture of both models
- Adding sensitivity to (token) frequency
- Changing more than one cell at once (N parameter)
- Adding conditioning for analogy (closer to Beniamine and Bonami, 2019)
- Simulating a tree-shaped evolution
- Comparing to the same measures on actual romance paradigms (Beniamine, Maiden, and Round, 2020)

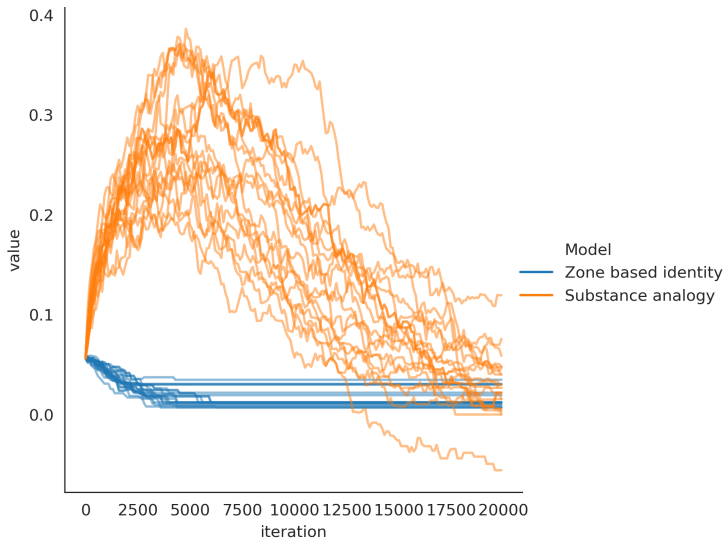
Conclusion

- The typical "affix conditional entropy" view of the PCFP can account for the development of inflection classes, but not morphemes
- Identity patterns should not be ignored, they are an important part of the implicative structure of paradigms, and can drive evolutionary processes.
- Analogy can mean many things: more work is necessary to understand exactly how it operates in inflection.
- This work is at the very start of a larger morpho-phonology project

Thank you for your attention.

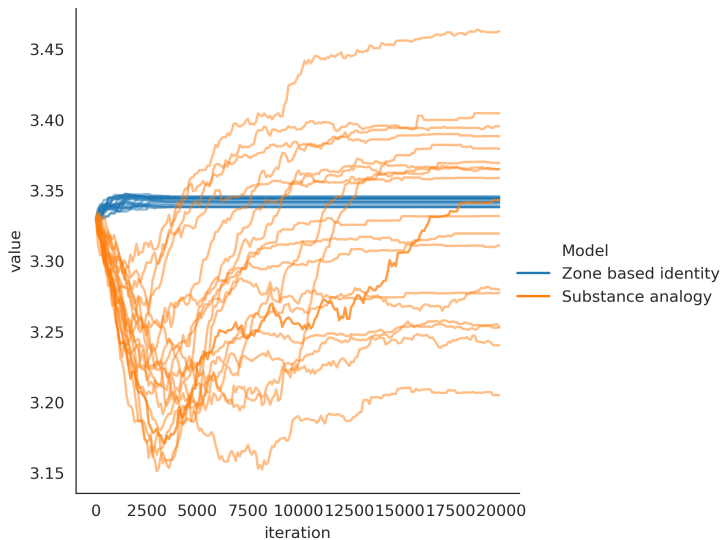
Paradigm entropy

U/L dataset



Cell entropy

U/L dataset



Morphome evolution

U/L dataset

