Beyond affix-stripping:

Generalisation and processing of 'pure morphology'

João Veríssimo

Potsdam Research Institute for Multilingualism

MoProc
24 June 2017

Core question

General research question

- What explains productivity?
- What *kind* of linguistic knowledge enables it?

Two broad approaches

Rule-based; dual-mechanism (e.g., Pinker, 1999)

- Some linguistic knowledge/processing involves:
- Rules (e.g., X $\rightarrow$ X*ed*$_{+\text{past}}$)
- Structured representations (e.g., [[walk][ed]])

Similarity-based approaches (Gonnerman et al., 2007)

- Morphological knowledge 'emerges' . . .
- . . . from regularities between form and meaning
- No rules
- No structured representations

Two domains for testing

Generalisation

- Context-free vs. context-sensitive operations
- Generalisation to nonce words (elicited production)

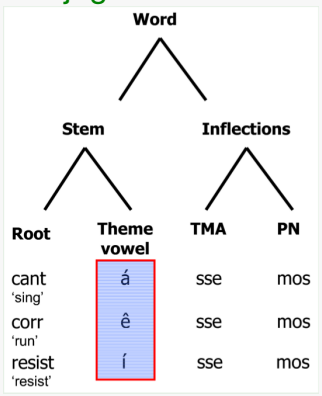
Decomposition

- Structured vs. 'undecomposed' representations
- Morphological priming (cross-modal, masked)

‘Pure’ morphology

Romance verb conjugations

Conjugation classes



- Theme vowels define three arbitrary classes
- Morphological, but not ‘meaning-bearing’

‘Pure’ morphology

Romance verb conjugations

[[[cant] a] $_{\text{stem}}$ va] $_{\text{past imp.}}$
[[[sorr] i] $_{\text{stem}}$ a] $_{\text{past imp.}}$

Conjugations as ‘pure’ morphology

- Theme vowels select verb endings
- Determine mappings between form and meaning
- “Irreducible morphological categories” (Aronoff, 1994)

‘Pure’ morphology

Romance verb conjugations

Striking discrepancy in productivity

- In Portuguese, Italian, etc. . . .
- 1st conj. welcomes novel words, borrowings, etc.
- 2nd and 3rd conjs. are seldom generalised

Example

- *to blog* ‘blogar’ (Port.), ‘bloggare’ (Ital.)
- [[[blog] $_{\text{root}}$ a] $_{\text{stem}}$ r]

General hypothesis

1st conj. stems

- Context-free rule: $X_{\text{root}} \rightarrow Xa_{\text{ stem}}$
- Generalised irrespective of phonological properties
- Constitute structured representations

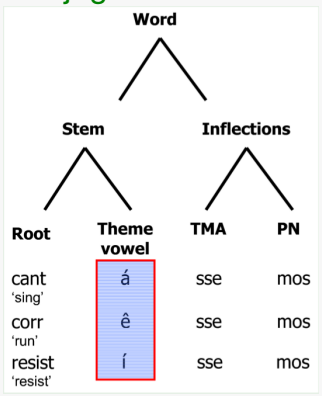
2nd and 3rd conj. stems

- ‘Exceptions’ to default stem-formation rule
- Generalisation is sensitive to phonology of root
- ‘Whole-stem’ representations

‘Pure’ morphology

Romance verb conjugations

Conjugation classes



- Theme vowels define three arbitrary classes
- Morphological, but not ‘meaning-bearing’

Our studies

Generalisation

- Elicited production (Port.)
- Computational simulations (Port.)
- Elicited production (Ital.)
- Reanalysis of Albright (2002) (Ital.)

Priming

- Cross-modal priming (Port.)
- Masked priming (Port.)

Generalisation of conjugations

Computational simulation

- Minimal Generalisation Learner (Albright, 2002)
- Input: Pairs of 1sg and Infinitive forms
- 1sg has no theme vowel
- Output: A set of phonological environments . . .
- . . . and corresponding reliability values for each class
- (e.g., in English past tense, *ing* is predictive of *i*→*a*)

Elicited production (pt)

Veríssimo & Clahsen (2014), *JML*

Method

- 54 native speakers of European Portuguese
- 78 novel verbs in the 1sg pres. ind. (*which does not display a theme vowel*)
- Participants had to fill a gap with an infinitive form (*which requires a theme vowel*)

Example

Quase sempre ***tureço*** *sozinho.*
Mas amanhã vou _____ *acompanhado.*
“I almost always **tureço** alone.
But tomorrow I will _____ someone.”

- Possible answers: *tureç**a**r*, *turec**e**r*, *turec**i**r*

Elicited production (pt)

Materials

MGL simulation

- Input: 3,117 Portuguese verbs
- 1sg to Infinitive

Construction of novel verbs

- 78 novel verbs created from MGL rules
- Spanning a wide range of reliability values
- Each novel verb is associated with 3 reliability values

Elicited production (pt)

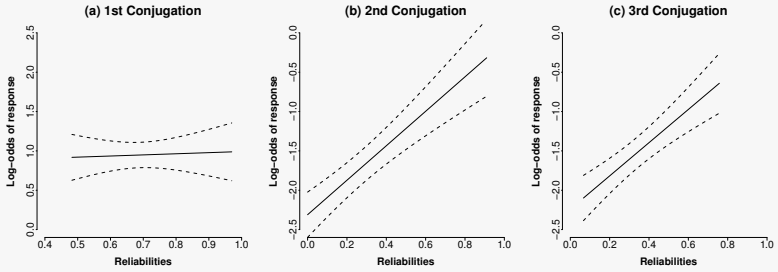
Results

- Three (weighted) regressions, predicting 1st, 2nd and 3rd conj. response log-odds
- Each w/ similarity to the 3 conjugations as predictors

	Responses (Log-odds)		
Predictors	*1st (-ar)*	*2nd (-er)*	*3rd (-ir)*
Reliab. 1st conj.	**.03**	−.04	.01
Reliab. 2nd conj.	−.65*	**.67***	−.10
Reliab. 3rd conj.	−.41*	−.16	**.58***

Elicited production (pt)

Results



Additional studies I

Model comparison (Veríssimo & Clahsen, 2014)

- Comparison of predicted proportions of responses in MGL vs. 'dual-mechanism' implementation
- Default Generalisation Learner (DGL)
- MGL underestimated 1st conj responses and overestimated 2nd and 3rd conj. responses
- DGL predictions for each of the three conjs. were statistically indistinghuishable from human responses

Additional studies II

Elicited production (Italian) (Veríssimo, in prep.)

- 35 native speakers
- 40 novel verbs (from Albright, 2002)
- 2nd conj. responses predicted by MGL reliabilities
- 1st conj. responses predicted by trade-off effects
- No significant effects for 3rd conj. responses

Additional studies III

Reanalysis of Albright (2002) (Veríssimo, in prep.)

- Acceptability judgements experiment (Italian)
- Ratings of 2nd and 3rd conj. forms were predicted by MGL reliability metric
- Ratings of 1st conj. forms were predicted by root well-formedness and trade-off effects

Generalisation in Romance

Discussion

- Generalisation of 1st conjugation in Romance languages is not sensitive to the phonological properties of novel roots (cf. Albright, 2002)
- 1st conj. generalised more widely than what would be predicted by the reliability metric
- Generalisation of 2nd and 3rd conjs. is based on phonological similarity

Results support a model that makes use of both context-free and similarity-based generalisations

Cross-modal priming

Veríssimo & Clahsen (2009), *Cognition*

Verb Type	*Prime Type*	*Primes*	*Targets*
1st Conjugation	Identity	limito	LIMITO 'I limit'
	Test	limitar	
	Control	desejar	
3rd Conjugation	Identity	adquiro	ADQUIRO 'I acquire'
	Test	adquirir	
	Control	investir	

Cross-modal priming
Method

Predictions

- 1st conj.: limit a r → limit o
- 3rd conj.: adquir i r → adquir o
- Larger stem → root priming for 1st conj.

Cross-modal priming
Method

Participants

- 57 native speakers of Portuguese (mean age: 26.1)

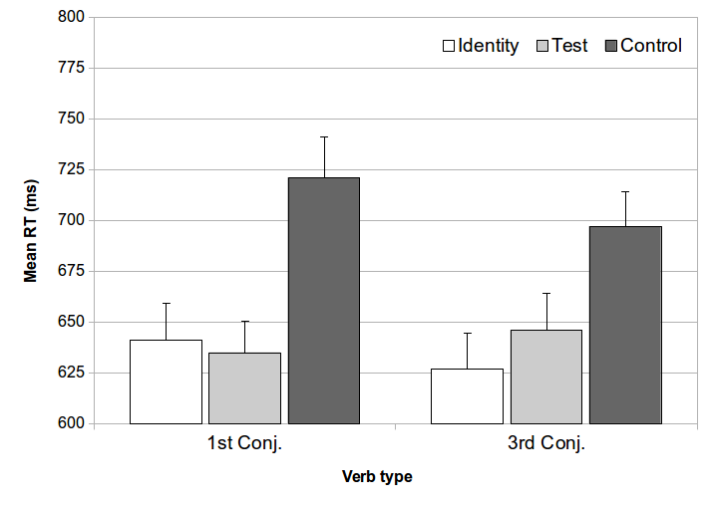
Materials

- 21 prime-target triplets in Verb Type condition
- Matched for lemma and form frequency, length, orthographic neighbours

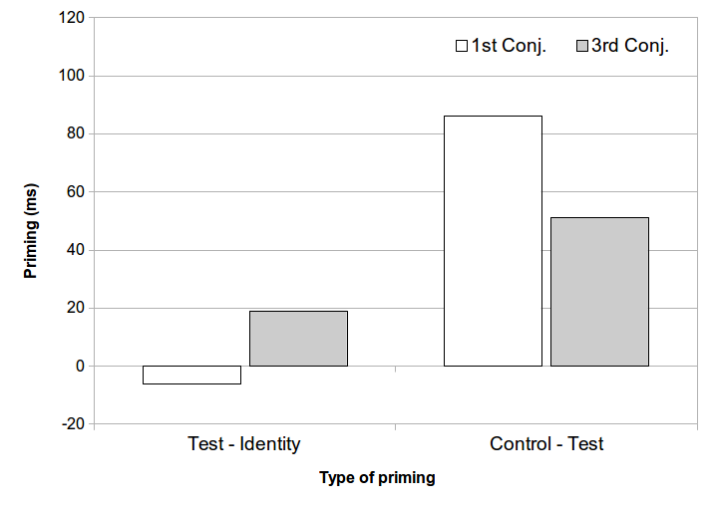
Procedure

- Cross-modal: auditory primes; visual targets
- Lexical decision task

Cross-modal priming
Results



Cross-modal priming
Results



Cross-modal priming
Discussion

- 1st conjugation produces a 'full priming' effect
- 3rd conjugation produces a 'partial priming' effect

Conclusion

- 1st conj. stems are structured ([root + tv])
- 3rd conj. stems are 'whole-stems', undecomposed

Masked priming
Veríssimo (in prep.)

Why masked priming?

- Arguably taps into 'access level' of representation
- Morphological effects that are less influenced by semantics

Masked priming
Method

Participants

- 60 native speakers of Portuguese (mean age: 26.0)

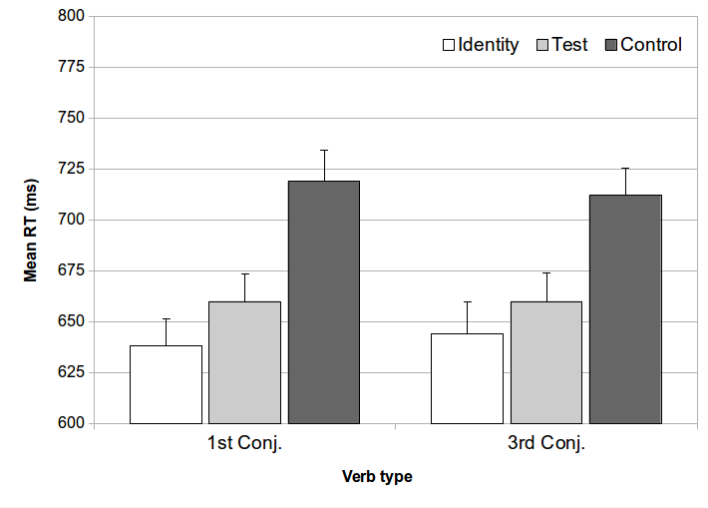
Materials

- Same 21 prime-target triplets in each condition
- Matched for lemma and form frequency, length, orthographic neighbours

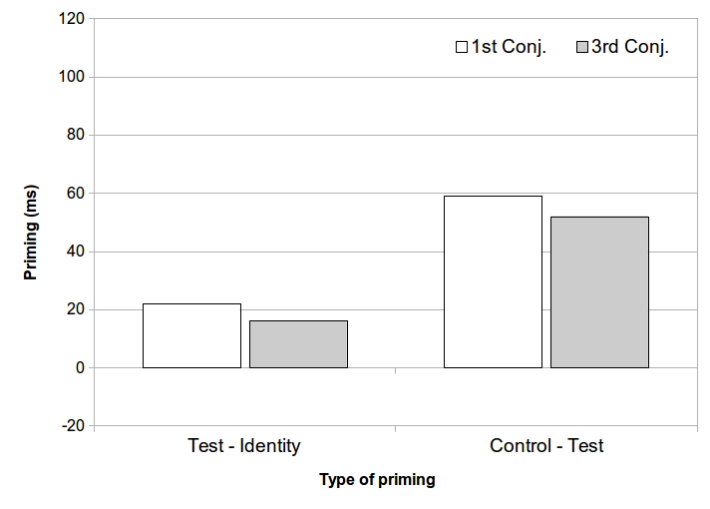
Procedure

- Masked priming: 67ms visual primes; visual targets
- Lexical decision task

Results

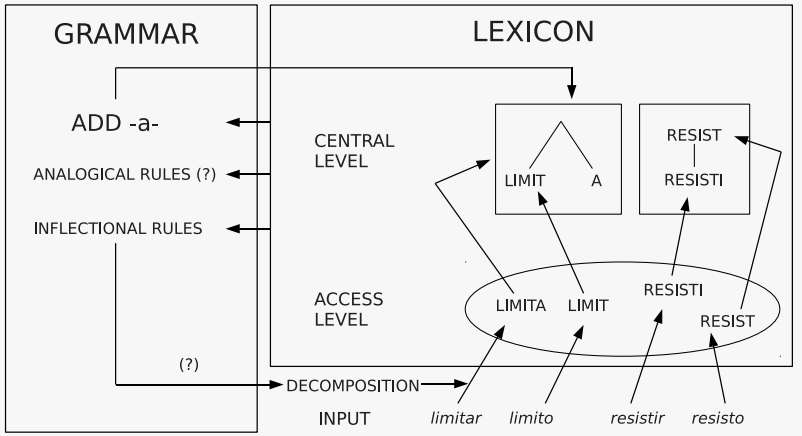


Results



Dual-mechanism model

Romance conjugations



Discussion

At least some morphological knowledge . . .

- . . . is couched in the form of context-free operations
- . . . is based on structured representations
- 'Non-default' morphology is particularly sensitive to (graded) phonological similarity

Discussion

Rule-based or stored stems?

- Storage of structured representations vs. 'whole-stem'?
- Format aligns with conjugation membership
- Accounts that postulate same representations for all classes fail to account for their productivity contrast
- Learning models could benefit from additional principles that 'partition' the conjugation space

Conclusion

For the ‘niche’ field of morphological processing . . .

- the study of more abstract morphological phenomena may contribute to solving long-standing theoretical controversies

Thank you!