From Latin to Romance: Computational Modeling of Syncretism

Tyler Lau*, Maria Polinsky†, and Jake Seaton†

*Department of Linguistics, University of California at Berkeley, USA
†Deparment of Linguistics, Harvard University, USA

Overview

Questions:

- What factors in Late Latin led to the heavy reshaping of the nominal system?
- What minimal information does a connectionist model need to predict syncretism in the correct direction?

Background:

- Analogy driven by factors such as *frequency*, *markedness*, and *morpheme length*. (Kurylowicz 1947, Bybee 1985, Albright 2008)
- From Latin to Romance
  - Declension: 5 > 3∼2 (I, II, (III)): frequency, sound change
  - Gender: 3 > 2 (M, F): sound change, contact
  - Case: 6 > 2∼1 (ACC, (NOM/GEN)): sound change, periphrastic constructions (preposition+ACC)
- Fate of the Neuter
  - N.SG had same endings as M.SG so many became M
  - N.PL ended in *-a* and as plural inanimates were seen as collectives, reinterpreted as F.SG: ex. Lat. *folia* 'leaves N.PL' > Sp. *hoja* 'leaf F.SG' (Herman 1967)
- Romanian has an ambigeneric system
  - "Neuter" class takes M morphology in singular and F in plural
  - Falls out from same principles as other Romance languages (with N.PL being reinterpreted as F.PL)
  - Many M's migrate to N class via analogy (Lat. *campus* ∼ *campi* 'field M' > Rom. *cîmp* (M.SG) ∼ *cîmpuri* (F.PL) , likely via *tempus* ∼ *tempora* 'time N')

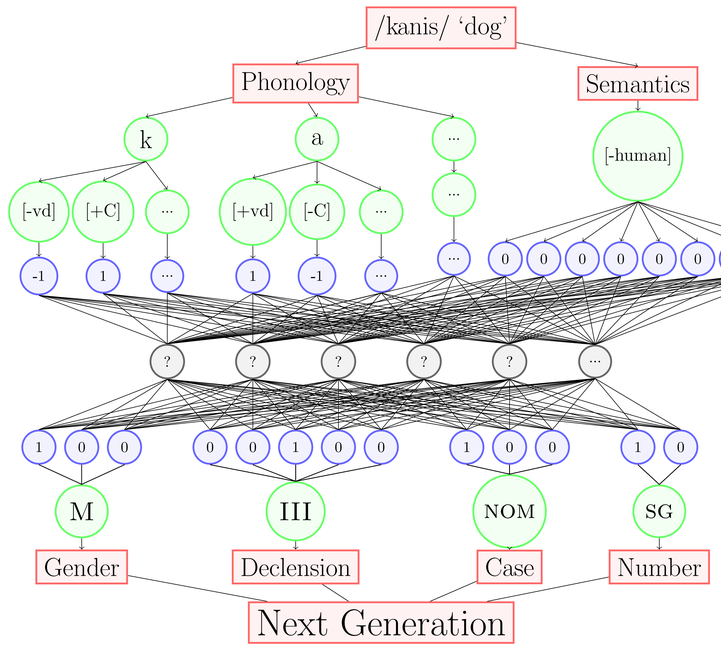
Objective: To use a connectionist simulation of generational learning providing minimal phonological and semantic information and see whether the changes that are actually attested in Romance can be reproduced

Latin Declension System

		I	II	IIIa	IIIb	IV	V
	Root	silv**a-**	ann**o-**	col**or-**	ign**i-**	lac**u-**	fid**e-**
	Gloss	'forest'	'year'	'color'	'fire'	'lake'	'faith'
Sg.	Nom.	silv**a**	ann**us**	color	ign**is**	lac**us**	fid**ēs**
	Gen.	silv**ae**	ann**ī**	colōr**is**	ign**is**	lac**ūs**	fid**eī**
	Acc.	silv**am**	ann**um**	colōr**em**	ign**em**	lac**um**	fid**em**
Pl.	Nom.	silv**ae**	ann**ī**	colōr**ēs**	ign**ēs**	lac**ūs**	fid**ēs**
	Gen.	silv**ārum**	ann**ōrum**	colōr**um**	ign**ium**	lac**ūm**	fid**ērum**
	Acc.	silv**ās**	ann**ōs**	colōr**ēs**	ign**īs**/ ign**ēs**	lac**ūs**	fid**ēs**

Figure 1: The Latin Declension Classes

Structure of the Connectionist Model



Input Layer: 454 Latin Vulgate nouns

Phonology (396 nodes = 6×6×11):
Each word maximally **6 syllables**
Each syllable maximally **6 phonemes** (CCVVCC)
Each phoneme coded for **11 features** 6 × 6 × 11

Semantics (8 nodes):
Human males activate initial 4 nodes
Human females activate final 4 nodes
Non-humans activate no nodes

Hidden Layer (30 nodes):
During training, hidden nodes use info from input and output to adjust connection weights. Hidden layer is the learner (Goldsmith & O'Brien 1995)

Output Layer (13 nodes):
Model had two parameters:
1) Case hierarchy? If yes → ACC (1,1,0) equidistant between NOM (1,0,0) and GEN (1,1,1)
2) Genitive drop? If yes → GEN node becomes unavailable output.
Output set of features given as expected output for following generation

Results

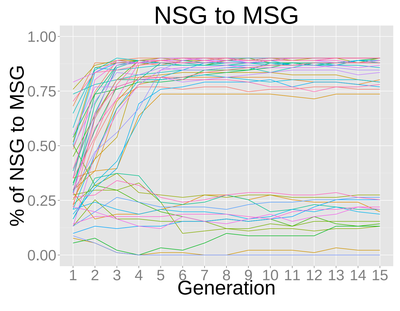


Figure 2: *With genitive drop, neuter singulars bifurcate–they either merge with masculines or drawing masculines to their class. The neuter plurals are drawn to the feminine singular class (as occurred in most Romance languages), but in a few trials merge with feminine plurals (as occurred in Romanian). Thus, the collapse of the neuter in most Romance languages and its bifurcation into an ambigeneric class in Romanian can be accounted for internally without appealing to contact effects.*

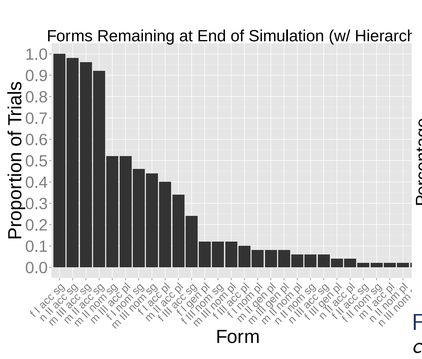


Figure 3: *With case hierarchy in play, the accusative is very robust, the genitive singular falls out completely, and the genitive plural survives in few cases. Without it, the genitive (both singular and plural) survive to a greater extent. The most robust forms are indeed widespread throughout Romance, including the well-known masculine -o, feminine -a, indeterminate -e/-∅, and plural -s endings.*

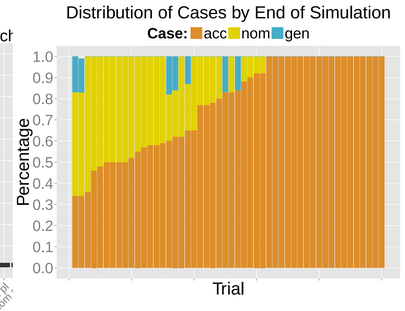


Figure 4: *With case hierarchy taken into consideration, the accusative becomes the dominant case in almost every simulation and the only case in almost half of them. The genitive survives in hardly more than 10% of trials (whereas without case hierarchy, the genitive remains in over 50% of trials)–thus, genitive drop may be predicted merely from phonology and semantics.*

Discussion

- With *phonology*, *frequency*, & *animacy semantics*
  - Declensions IV & V fall out in *every* simulation
- With *case hierarchy* added, final forms converge more
  - Genitive singular drops out *completely*
  - Genitive plural hardly survives (only example is oblique 3PL pronoun–Fr. *leur*, It. *loro*, Rom. *lor*)
  - Forms remaining in ≥90% of simulations
    - *-am* > *-a* F.SG ending in all Romance (> *-e* in Fr.)
    - *-um* > *-u* M.SG ending in all of Romance (> *-o* in Sp., It. etc.)
    - *-em* > *-e* SG ending for M/F nouns in all of Romance
  - Forms remaining in 25-90% of simulations
    - *-∅* SG ending for M/F nouns in all of Romance
    - *-ēs* PL ending in western Romance, maybe > *-i* in eastern
    - *-ōs* M.PL ending in western Romance, maybe > *-i* in eastern
    - *-ās* F.PL ending in western Romance, maybe > *-e* in eastern
    - M/F.NOM.SG *-us* & *-as*: in E-Romance., final *-s* falls out; in W-Romance, NOM persists in older Sp. & Fr.
  - Case system converges to accusative in almost half (as in western Romance), eastern Romance shows alternate history where nominative plural may have survived (see D'hulst (2006) on Romance plurals)
  - Neuter rarely survives–when it stays, a sizeable chunk of the masculine class migrates over (as in Romanian)
- With *genitive* dropped, N.SG > M.SG, N.PL > F.SG
  - Otherwise, N.PL.GEN > M.PL because of phonology
  - Supports popularity of periphrastic construction view

References

Albright, A. 2008. Explaining Universal Tendencies and Language Particulars in Analogical Change. In *Linguistics Universals and Language Change*, 144-181. Bybee, J. 1985. *Morphology: A Study of the Relation between Meaning and Form.* D'hulst, Y. 2006. Romance Plurals. *Lingua* 116, 1303-1329. Goldsmith, J. & O'Brien, J. 1995. Grammar within a Neural Net. In *The Reality of Linguistic Rules*, 95-113. Herman, J. 1967. *Le Latin Vulgaire.* Kuryłowicz, J. 1947. The Nature of the so-called Analogical Processes. *Diachronica* 12 (1), 113-145.

Acknowledgements

Many thanks to Kevin Ryan, James Kirby, Andrew Garrett, Terry Regier, Mairi McLaughlin, and Yang Xu for comments and guidance throughout the project, to Ezra Van Everbroeck for providing the code for the simulation in Polinsky and Van Everbroeck (2003), and to Edwin Ko for consultation on data visualization.

Contact Information

- Tyler Lau: tylerlau@berkeley.edu
- Maria Polinsky: polinsky@fas.harvard.edu
- Jake Seaton: jseaton@college.harvard.edu