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

- 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?
- Analogy driven by factors such as *frequency*, *markedness*, and *morpheme length*. (Kuryłowicz 1947, Bybee 1985, Albright 2008)
- Changes in Romance have been attributed not only to sound change, but also to contact
- We aim 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
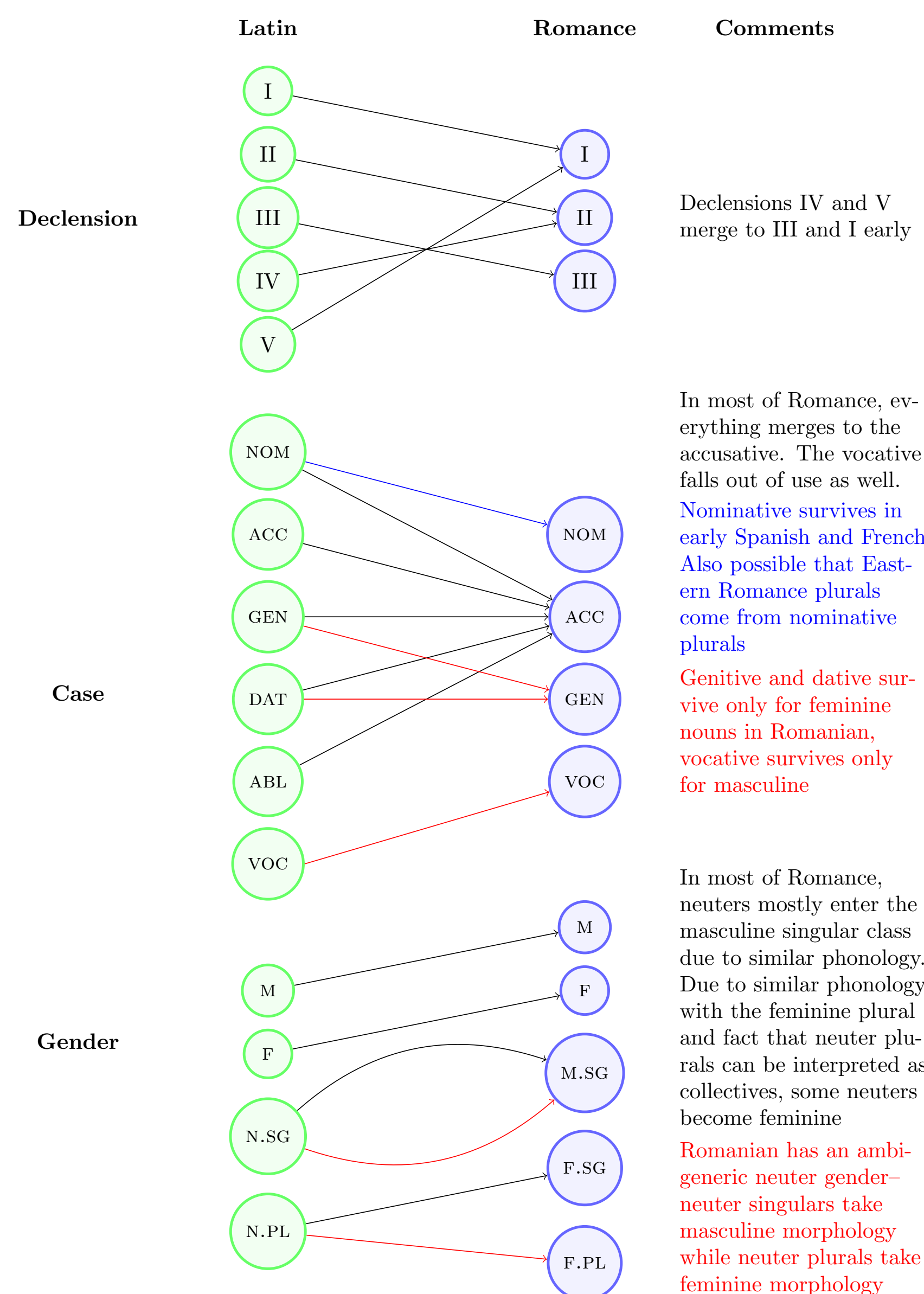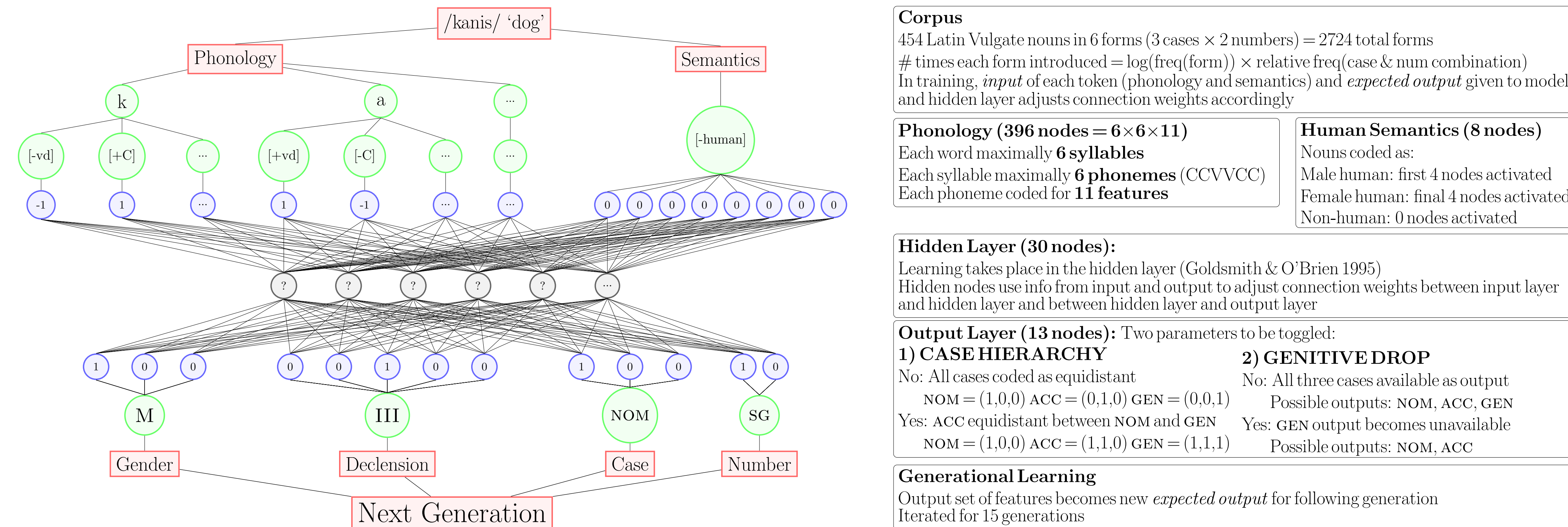

Latin Declension System

	I	II	IIIa	IIIb	IV	V
Root	silva-	anno-	color-	igni-	lacu-	fide-
Gloss	'forest'	'year'	'color'	'fire'	'lake'	'faith'
Sg. Nom.	silva	annus	color	ignis	lacus	fidēs
Sg. Gen.	silvae	annī	colōris	ignis	lacūs	fideī
Sg. Acc.	silvam	annum	colōrem	ignem	lacum	fidem
Pl. Nom.	annī	colōrēs	ignēs	lacūs	fidēs	
Pl. Gen.	silvārum	annōrum	colōrum	ignium	lacūm	fidērum
Pl. Acc.	silvās	annōs	colōrēs	ignīs/ignēs	lacūs	fidēs

Figure 1: The Latin Declension Classes

Structure of the Connectionist Model

Corpus
454 Latin Vulgate nouns in 6 forms (3 cases × 2 numbers) = 2724 total forms
times each form introduced = log(freq(form)) × relative freq(case & num combination)
In training, *input* of each token (phonology and semantics) and *expected output* given to model and hidden layer adjusts connection weights accordingly

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

Human Semantics (8 nodes)
Nouns coded as:
Male human: first 4 nodes activated
Female human: final 4 nodes activated
Non-human: 0 nodes activated

Hidden Layer (30 nodes):
Learning takes place in the hidden layer (Goldsmith & O'Brien 1995)
Hidden nodes use info from input and output to adjust connection weights between input layer and hidden layer and between hidden layer and output layer

Output Layer (13 nodes): Two parameters to be toggled:
1) CASE HIERARCHY
No: All cases coded as equidistant
    NOM = (1,0,0) ACC = (0,1,0) GEN = (0,0,1)
Yes: ACC equidistant between NOM and GEN
    NOM = (1,0,0) ACC = (1,1,0) GEN = (1,1,1)
2) GENITIVE DROP
No: All three cases available as output
    Possible outputs: NOM, ACC, GEN
Yes: GEN output becomes unavailable
    Possible outputs: NOM, ACC

Generational Learning
Output set of features becomes new *expected output* for following generation
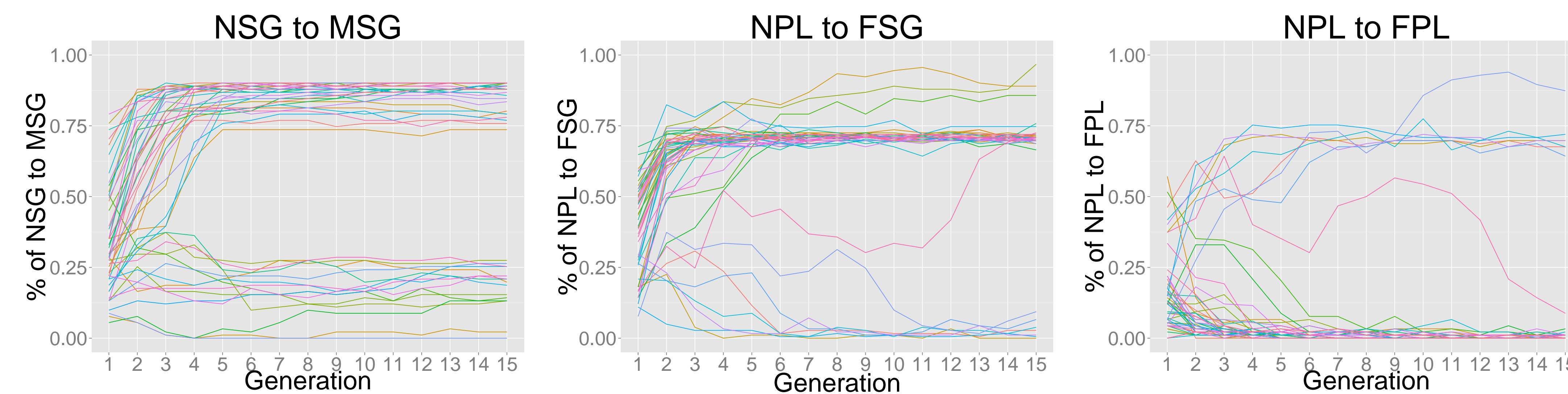Iterated for 15 generations

Results

Figure 2: *With genitive drop, neuter singulars bifurcate–they either merge with masculines or draw masculines to their class (see Figure 6).*

Figure 3: *With genitive drop, neuter plurals almost consistently migrate to the feminine singular class due to phonological similarity alone.*

Figure 4: *Comparing with Figure 3, it is notable that in those trials where neuter plurals do **not** migrate to the feminine singular class, they migrate to the feminine plural.*
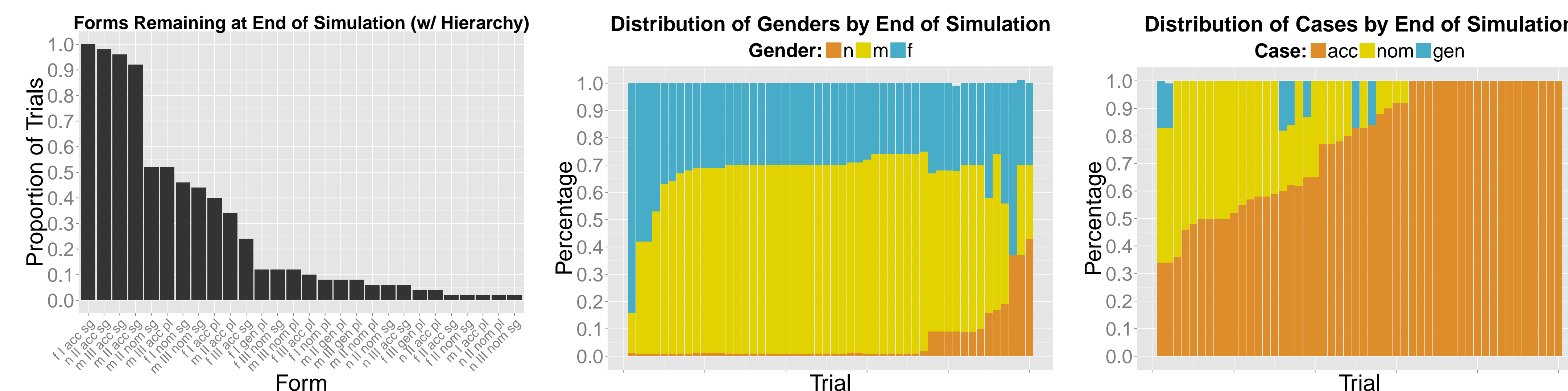

Figure 5: *With case hierarchy in play, accusative is very robust and only the genitive plural survives in some trials. Without it, the genitive (both singular and plural) survive to a greater extent.*

Figure 6: *In most of the trials, the neuter falls out. In the cases where it is more robust, masculine nouns migrate to the neuter class and the total proportion of masculine+neuter nouns remains approximately equal.*

Figure 7: *With case hierarchy taken into consideration, the accusative becomes the dominant case in most trials and the only case in almost half of them. The genitive survives in hardly more than 10% of trials (without case hierarchy, the genitive remains in over 50% of trials).*

Discussion

- With *phonology*, *frequency*, & *human 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 in history is oblique 3PL pronoun–Fr. *leur*, It. *loro*)
  - 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.
- Accusative is most robust form–in history, acts as base form in most of modern Romance. Plurals in some languages *may* be from nominative (D'hulst 2006)
- With *genitive* dropped, two notable outcomes for neuter
  - ❶ N.SG > M.SG | N.PL > F.SG (most of Romance)
  - ❷ M.SG > N.SG | N.PL > F.PL (Romanian system)
- Taking into account these minimal factors, simulation offers a rather accurate history of syncretism and trends that occurred on way to modern Romance languages

References

Albright, A. 2008. Explaining Universal Tendencies and Language Particulars in Analogical Change, 144-181. In *Linguistics Universals and Language Change*. 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

Contact Information

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