On some parallels between defective and normal inflection

Andrea D. Sims
The Ohio State University
sims.120@osu.edu

Quantitative Measures in Morphology & Morphological Development Workshop
University of California, San Diego
January 16-17, 2011

What is inflectional defectiveness?

i θálassa 'the sea'	**singular**	**plural**
nominative	θálassa	θálasses
accusative	θálassa	θálasses
genitive	θálassas	θalassón
vocative	θálassa	θálasses

i kopéla 'the girl'	**singular**	**plural**
nominative	kopéla	kopéles
accusative	kopéla	kopéles
genitive	kopélas	*
vocative	kopéla	kopéles

Table 1: Defectiveness in the Modern Greek genitive plural

missing form = paradigmatic gap

What is inflectional defectiveness?

sprosit' 'to ask'	**singular**	**plural**
1st	sprošu	sprosim
2nd	sprosiš'	sprosite
3rd	sprosit	sprosjat

ubedit' 'to convince'	**singular**	**plural**
1st	*	ubedim
2nd	ubediš'	ubedite
3rd	ubedit	ubedjat

Table 2: Defectiveness in the Russian non-past 1sg

Classically, gaps are seen to contradict the productive nature of inflectional morphology, fly in the face of speakers' tendency to generalize, and represent ad hoc exceptions to general grammatical principles (see, e.g., Halle 1973).

So how weird *are* gaps?

- Premise: If irreducible patterns of defectiveness are, in some sense, 'morphological objects' (i.e. lexically-specific information about morphological structure), we can compare the properties of gaps to the properties of 'normal' (i.e. non-defective) word-forms.

- Question: Do paradigmatic gaps exhibit properties that are typical of inflectional formatives?

- Preview: Perhaps surprisingly, paradigmatic gaps can display many of the same properties as inflectional morphology generally.

Some properties of inflectional formatives

- ~~Part of a relatively small closed system~~
- ~~Shape determined by properties of the base~~
- High degree of stem selection
  - Incl. lexical specificity
- Participate in blocking relations
- ~~Prone to lexicalization~~
- Vary in degree of productivity
- Sensitive to paradigmatic structure
- Sensitive to the structure of lexical neighborhood
  - Analogical spreading of defectiveness

1. 'Productivity' of defectiveness?

What we are looking for:

greater 'market share' among the least frequent lexemes

Reminder: Paradigmatic gaps in Greek

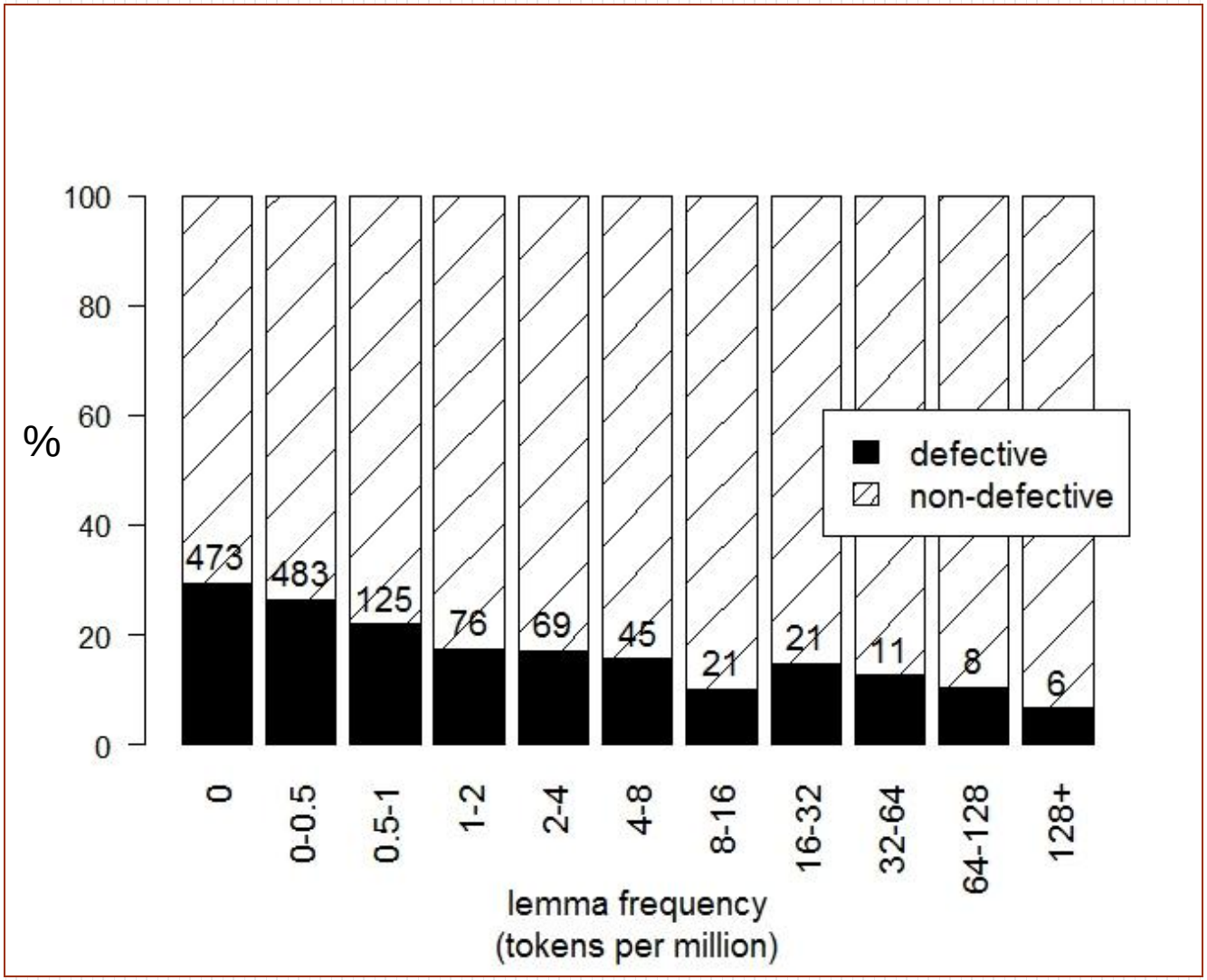
i θálassa 'the sea'	**singular**	**plural**
nominative	θálassa	θálasses
accusative	θálassa	θálasses
genitive	θálassas	θalassón
vocative	θálassa	θálasses

i kopéla 'the girl'	**singular**	**plural**
nominative	kopéla	kopéles
accusative	kopéla	kopéles
genitive	kopélas	*
vocative	kopéla	kopéles

Table 1: Defectiveness in the Modern Greek genitive plural

- According to online version of the *Lexiko tis koinis neoellinikis* (1998), 1560 nouns are defective in the genitive plural.
- Not evenly distributed across inflection classes

Distribution of defective lexemes



Low-frequency lexemes are more likely to be defective than are high frequency lexemes

1. 'Productivity' of defectiveness?

2. Sensitive to paradigmatic structure?

What we are looking for:

generalizations that require reference to multiple cells in the paradigm

Paradigmatic structure

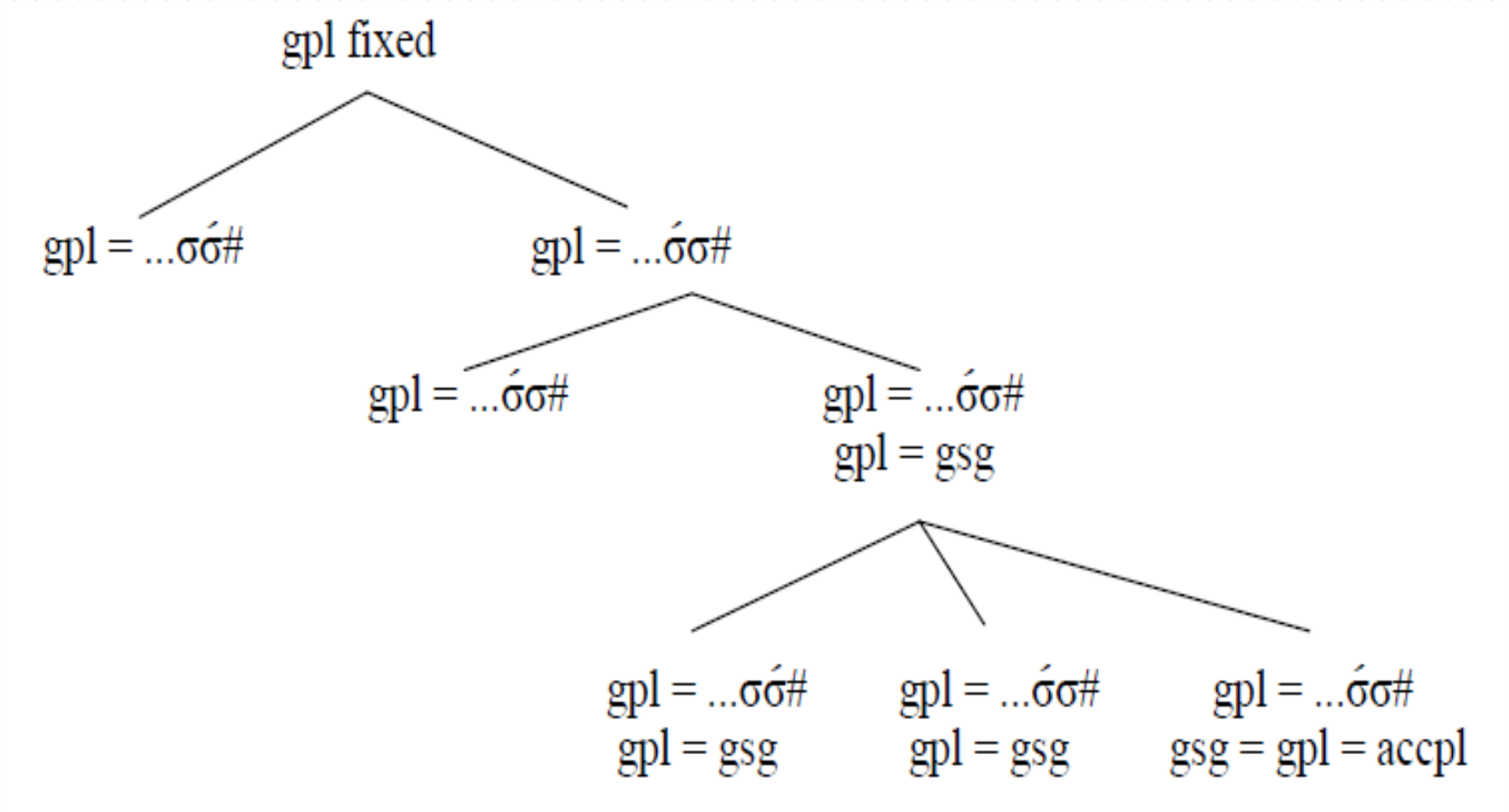
- Stump (2010): Defectiveness and syncretism interact in Sanskrit such that one may override the other
  - See also Hansson (1999); Gaps in Icelandic imperatives can be construed as an instance of syncretism overriding defectiveness

- Paradigmatic coherence: Form-level implicational relations holding between cells of a paradigm (Ackerman et al. 2009; Bonami and Boyé (2002); Brown et al. 1996; Finkel and Stump 2007, 2009)
  - See Boyé and Cabredo Hofherr (2010) for how implicational relations among stems helps make sense of paradigmatic gaps in Spanish and French

Greek nominal inflectional formatives

Singular Formatives				Plural Formatives		
NOM	GEN	ACC		NOM	GEN	ACC
-Ø	-s	-Ø		-es	-on	-es
-s	-Ø	-Ø		-is	-on	-is
-os	-u	-o		-i	-on	-us
-o	-u	-o		-i	-on	-i
-Ø	-u	-Ø		-a	-on	-a
-os	-us	-os		-is	-eon	-is
-as	-os	-a		-des	-don	-des
-Ø	-tos	-Ø		-ides	-idon	-ides
-s	-tos	-s		-ta	-ton	-ta
-ma	-matos	-ma		-mata	-maton	-mata
-mo	-matos	-mo		-antes	-anton	-antes

Stress hierarchy



(Lack of) correspondence between singular, plural and stress formatives

	'force'	'mother'	'greengrocer'	'tourist'	'father'
NOM.SG.	*đínami*	*mamá*	*manávi-s*	*turísta-s*	*patéra-s*
ACC.SG.	*đínami*	*mamá*	*manávi*	*turísta*	*patéra*
GEN.SG.	*đínami-s*	*mamá-s*	*manávi*	*turísta*	*patéra*
NOM.PL.	*đínami-s*	*mamá-đes*	*manávi-đes*	*turíst-es*	*patér-es*
ACC.PL.	*đínami-s*	*mamá-đes*	*manávi-đes*	*turíst-es*	*patér-es*
GEN.PL.	*đínam-eon*	*mamá-đon*	*manávi-đon*	*turíst-ón*	*patér-on*

A 'bit' of introduction to entropy

$$H(X) = -\sum_{x \in X} p(x) \log_2 p(x)$$

- H(X) is the entropy of the morphosyntactic property set (MSPS; e.g. genitive plural)
- Function of average surprisal
- Measured in bits
  - HIGH value for H(X) means MORE UNCERTAINTY associated with the MSPS
  - LOW value for H(X) means LESS UNCERTAINTY associated with MSPS

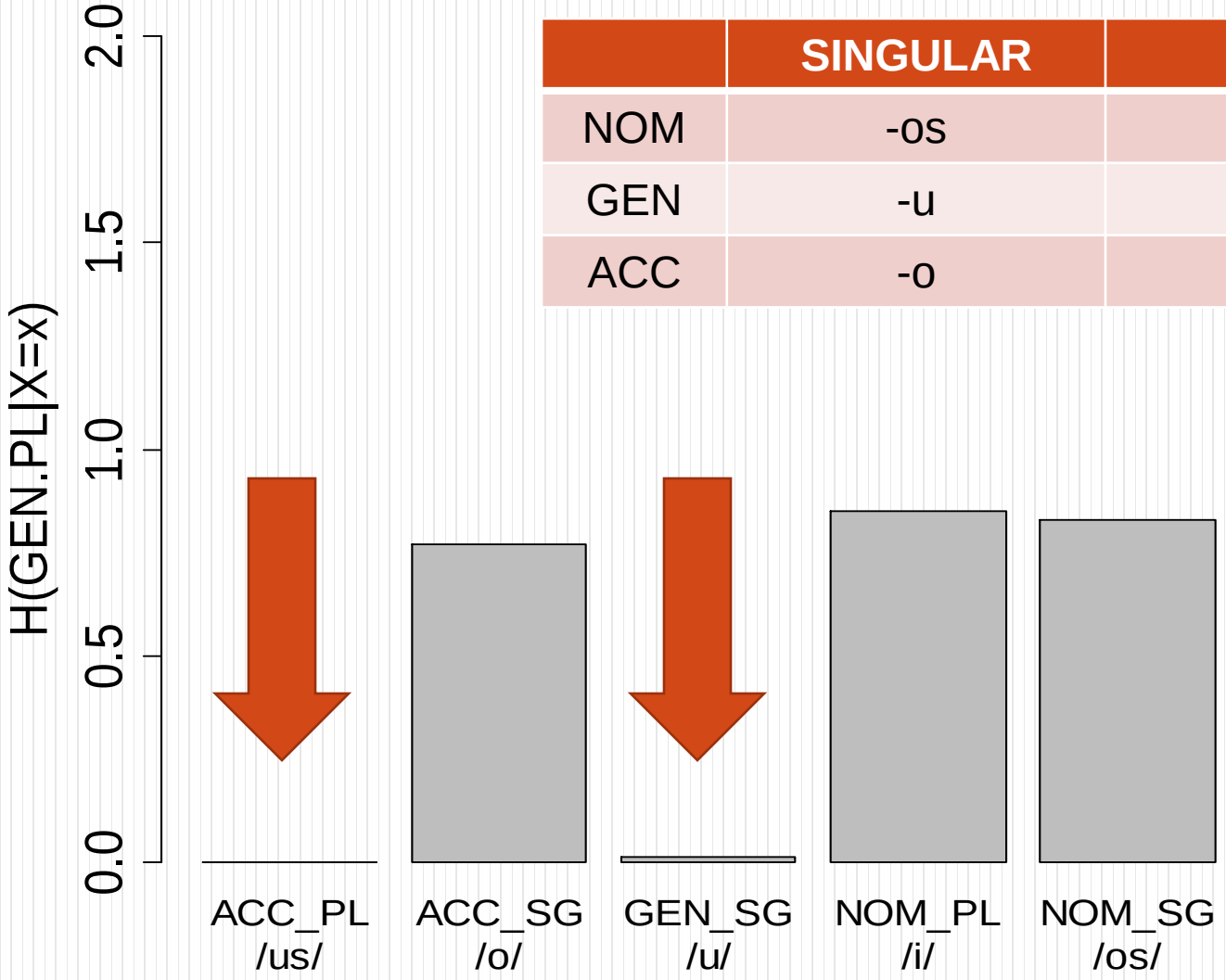
A 'bit' of introduction to entropy

$$H(Y|X=x) = -\sum_{y \in Y} p(y|x) \log_2 p(y|x)$$

- H(Y|X=x) is the conditional entropy of the MSPS Y, given a particular value x, belonging to morphosyntactic property set X
- For instance, how much uncertainty is associated with the form of the word occupying the genitive plural cell, given that the nominative singular has the form *-os*?

Class O17-O20, O34-O36

	SINGULAR	PLURAL
NOM	-os	-i
GEN	-u	??
ACC	-o	-us

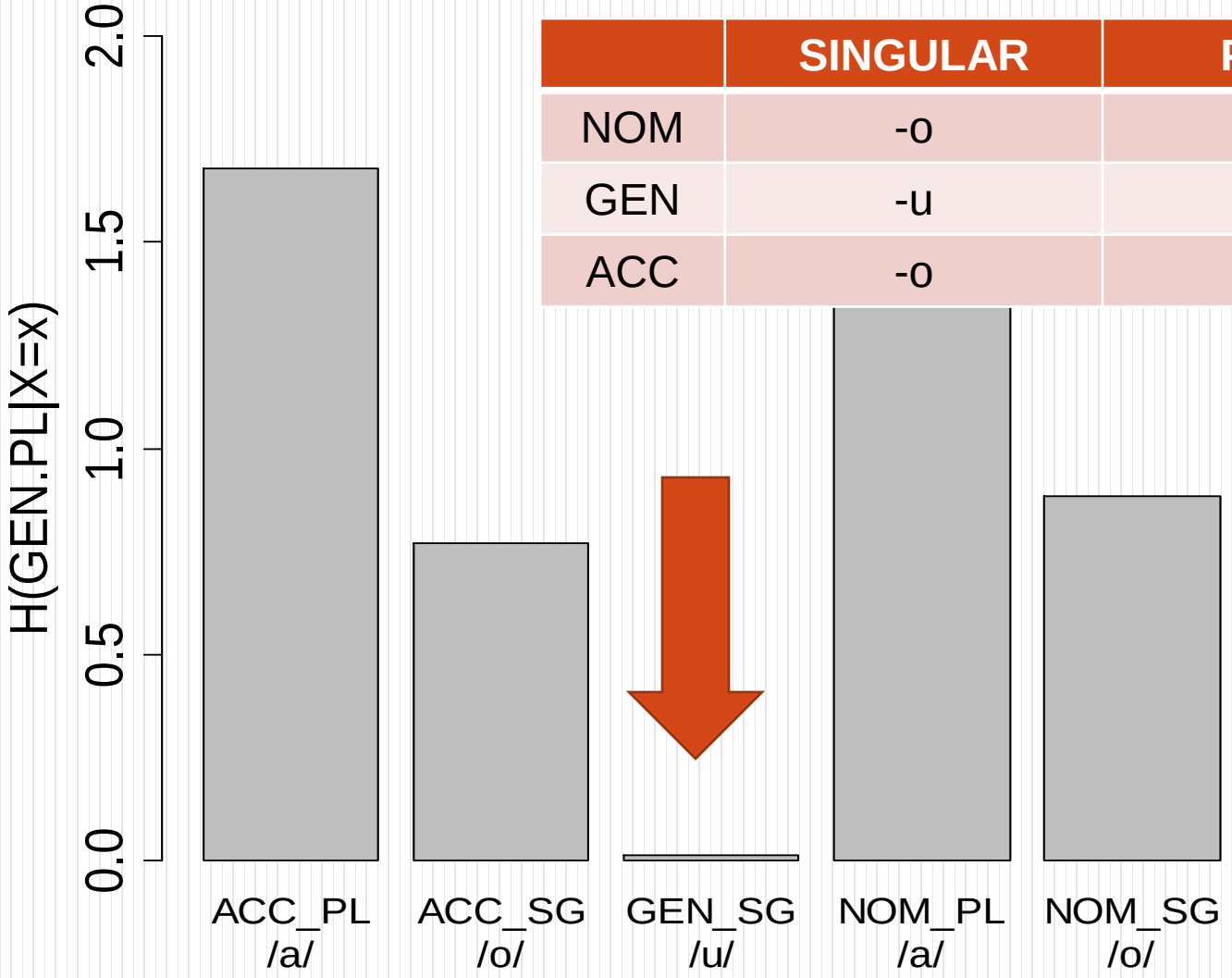


X={ACC_PL,…}, x={/us/,…}

Reading graph:
Low entropy value = low level of uncertainty associated with genitive plural form (i.e. genitive plural form is more predictable)

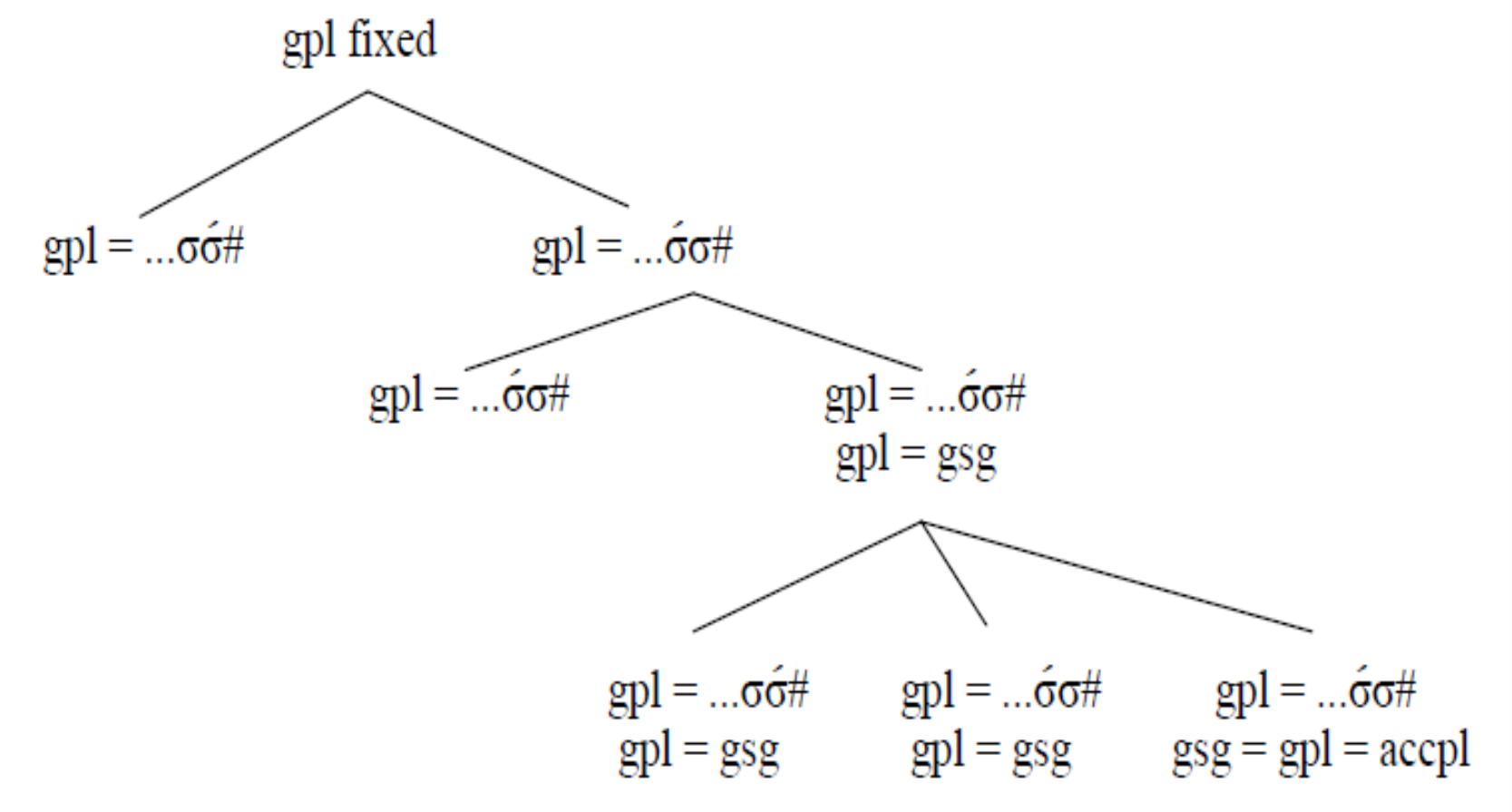
Class O39-O42

	SINGULAR	PLURAL
NOM	-o	-a
GEN	-u	??
ACC	-o	-a



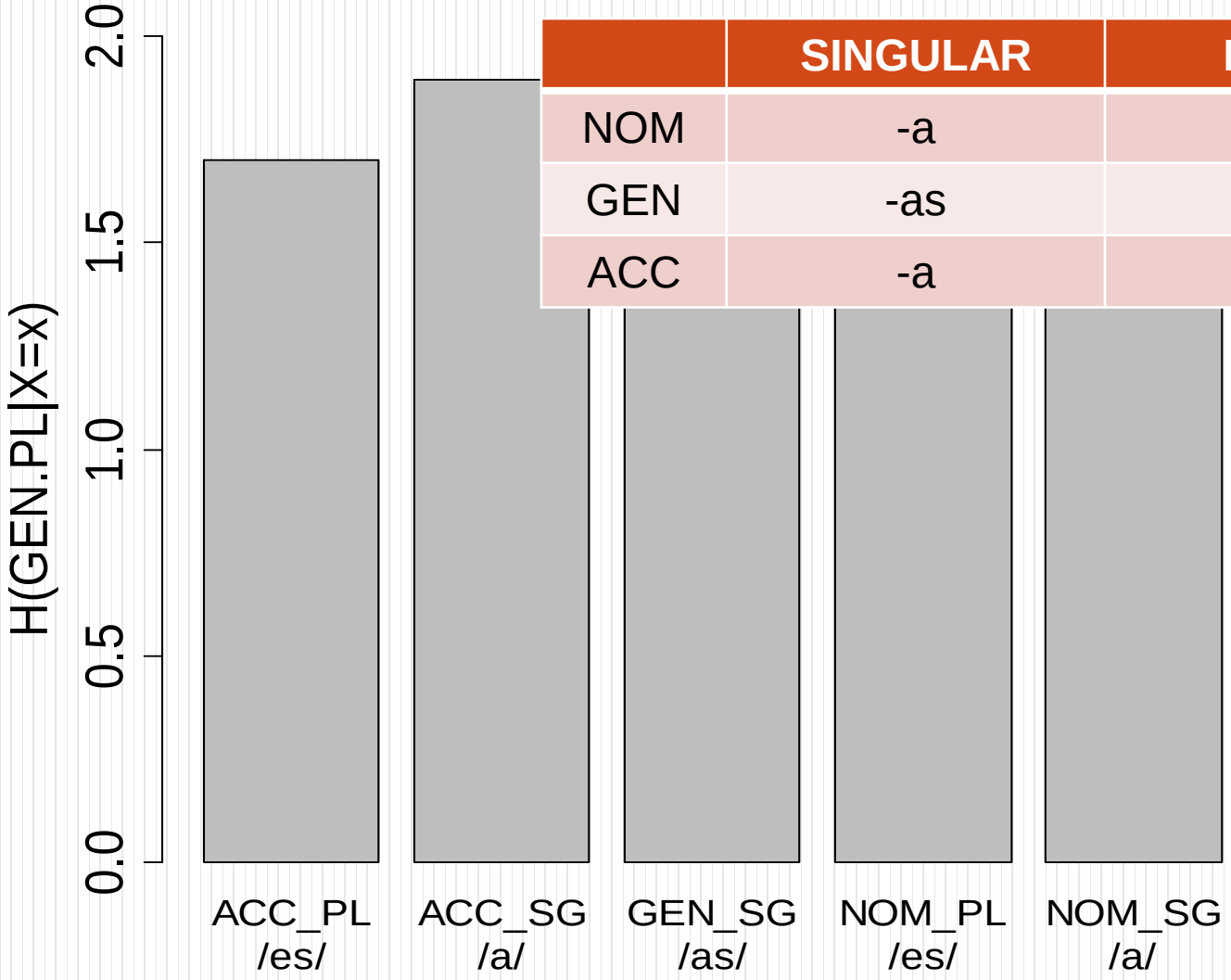
Reading graph:
Low entropy value = low level of uncertainty associated with genitive plural form (i.e. genitive plural form is more predictable)

Reminder: Stress hierarchy



Class O24-O28, O45

	SINGULAR	PLURAL
NOM	-a	-es
GEN	-as	??
ACC	-a	-es



X={ACC_PL,...}, x={/es/,...}

This is the class containing (almost) all of the defective verbs!

Conclusion from Greek entropy data

- Uncertainty associated with genitive plural form (and with other forms given genitive plural) is (historically) connected to existence of paradigmatic gaps.

- This suggests that paradigmatic gaps in Greek are sensitive to the implicational relations holding among cells in a paradigm.
  - Parallels to, e.g., analogical extension

2. Sensitive to paradigmatic structure?

3. Sensitive to the structure of the lexicon?

What we are looking for:

evidence that paradigmatic gaps behave similarly to lexical gangs

Reminder: Gaps in Russian verbs

sprosit' 'to ask'	**singular**	**plural**
1st	sprošu	sprosim
2nd	sprosiš'	sprosite
3rd	sprosit	sprosjat

ubedit' 'to convince'	**singular**	**plural**
1st	*	ubedim
2nd	ubediš'	ubedite
3rd	ubedit	ubedjat

- The challenge: Under what conditions can paradigmatic gaps lose their original motivating factors and still persist? What (type of) information must be available to a speaker for lexicalized defectiveness to be learnable?

More about the Russian gaps

- Key observation: All of the defective lexemes belong to morphological subclass of 2nd conjugation dental stems
  - See Baerman (2008) for discussion of historical causes
- However, defective lexemes are in the minority even within that subclass
- Moreover, many of the defective lexemes are quite infrequent

Russian lexemes: dental class	**stem-final /d^j/**	**stem-final /t^j/**	**stem-final /z^j/**	**stem-final /s^j/**	**stem-final /st^j/**
gaps / all lexemes (RNC)	13.3% (19/143)	12.4% (14/118)	11.9% (5/42)	4.8% (3/62)	4.3% (2/47)

The challenge

- How do speakers learn that a given verb is defective when...
  - ... the number of encountered examples of the lexeme is small
    - Suggests that we need more than statistical inference from lack of attestation
  - ... and well-formedness is always more likely within the class than is defectiveness?
    - And we would expect, if anything, for there to be a bias towards speakers eliminating the gaps

A hypothesis

- It's about expectations. Paradigmatic gaps are usually considered violations of expected morphological behavior. Is this correct?
  - (Non-defective) low frequency lexemes rely on their neighbors.
  - Suggests the possibility that for some lexemes, "defectiveness" is expected – if a sufficient number of the neighbors are also defective.
    - Self-reinforcement of morphophon.-defined clusters of gaps.
- Two ways to learn gaps
  - (In the absence of synchronic motivation), learning a gap involves estimating the (near-zero) probability of a given combination of lexeme and inflectional property set being used.
    - Word-specific learning for highly frequent lexemes
    - Analogically-driven learning from lexical neighbors for lower frequency lexemes
- Observed data and morphological class/neighbors may individually be insufficient, but together can they pick out the correct subset of lexemes?

For example...

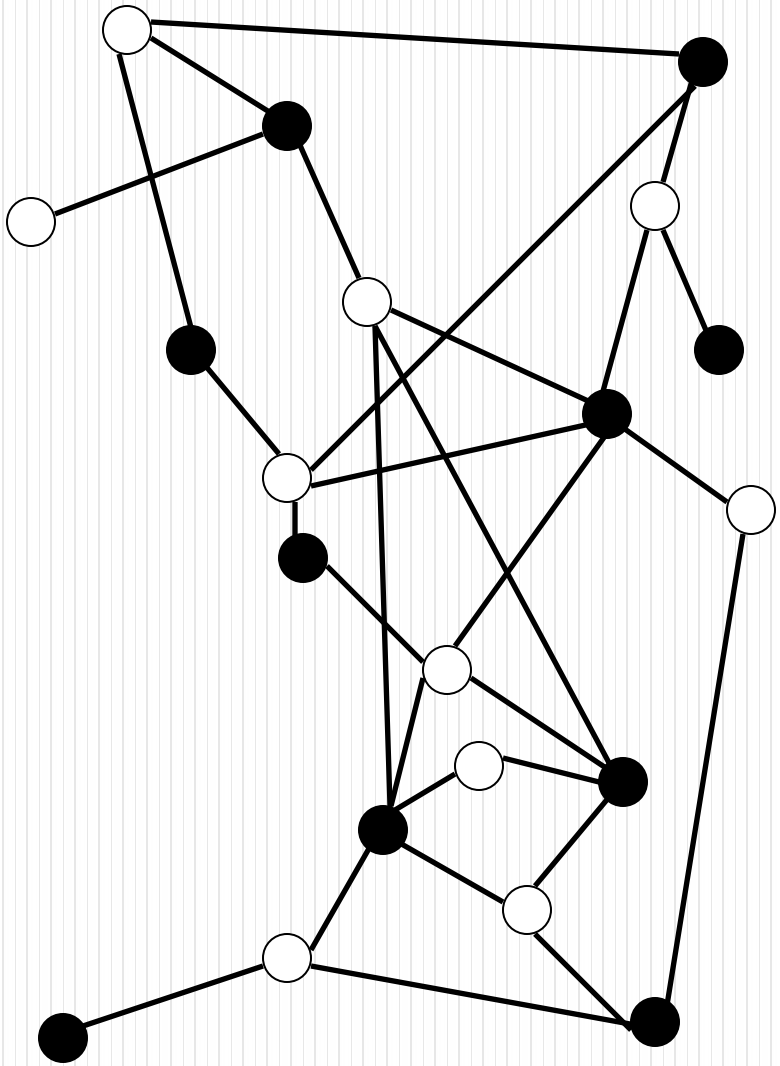
ubedit' *'convince'*	Raw #	Relative freq	'Normal' lexemes
1SG	1	0.2%	12.9%
2SG	53	11.7%	7.5%
3SG	210	46.4%	38.7%
1PL	27	6%	10.6%
2PL	71	15.7	9.7%
3PL	91	20.1	20.6%
Sum	453	100	100

– If learner hears many UBEDIT' tokens, but no or few tokens of UBEDIT'+1SG, infers that relative absence is a property of UBEDIT'

– If learner hears few tokens (e.g. KUDESIT' 'do magic'), distribution of lexical neighbors is more influential

A computational simulation

- Adults talk (100,000 nouns each), children listen
- End of generational cycle: adults die off, children learn grammar, mature, reproduce
- Speech of new adults based on the grammar that they learned
- 10 generations
- 50 adults and 50 children per generation
- Each child connected to 10 adults on average (random network)
- First generation seeded by sampling from Russian National Corpus



Evaluating the model

- Evaluation question
  - Does the number of gaps remain (relatively) constant for multiple generations?
- Conditions
  - Two types of analogical influence from lexical neighborhood: unweighted vs. morphophonologically-weighted
  - Four levels of analogical influence

Morphophonological similarity metric

- Weighting metric: In MP condition, weight = {1, 0.67, 0.33, 0} depending on phonological feature distance of stem-final consonant, comparing target to neighbor
  - e.g., both /dʲ/ → 1
  - e.g., one /tʲ/, one /dʲ/ → 0.67 (one feature difference)
- Gap criteria
  - Remove low sampling : raw lemma frequency > 37 tokens in output of model at each generation
  - No impersonal verbs: 3sg+3pl < 85% of relative freq
  - 1sg < 2% relative frequency
    - (2% = valley of bimodal distribution)

target: *ubedit'*

1) Lexical nbhd

balamutit'
besit'sja
bespokoit'
božit'sja
brodit'
vesit'
vzvintit'
vkatit'
vykrasit'
grozit'sja
dobavit'
...
pobedit'
sekonomit'
sudit'
jutit'sja
javit'sja

2) Average relative freq for lexical neighborhood

neighborhood	w	1s	2s	3s	1p	2p	3p
balamutit'	**2/3**	7.1	7.1	57.1	0	0	28.6
brodit'	**1**	9.2	3.7	48.1	3.3	1.8	33.8
vykrasit'	**1/3**	42.8	0	28.6	7.1	7.1	14.3
pobedit'	**1**	0.1	4.5	61.9	16.8	1.5	15.1
javit'sja	**1/3**	5.1	2.8	69.9	1.3	3.5	17.3
Average relative freq (lexical neighborhood)		12.9	3.6	53.1	5.7	2.8	21.8

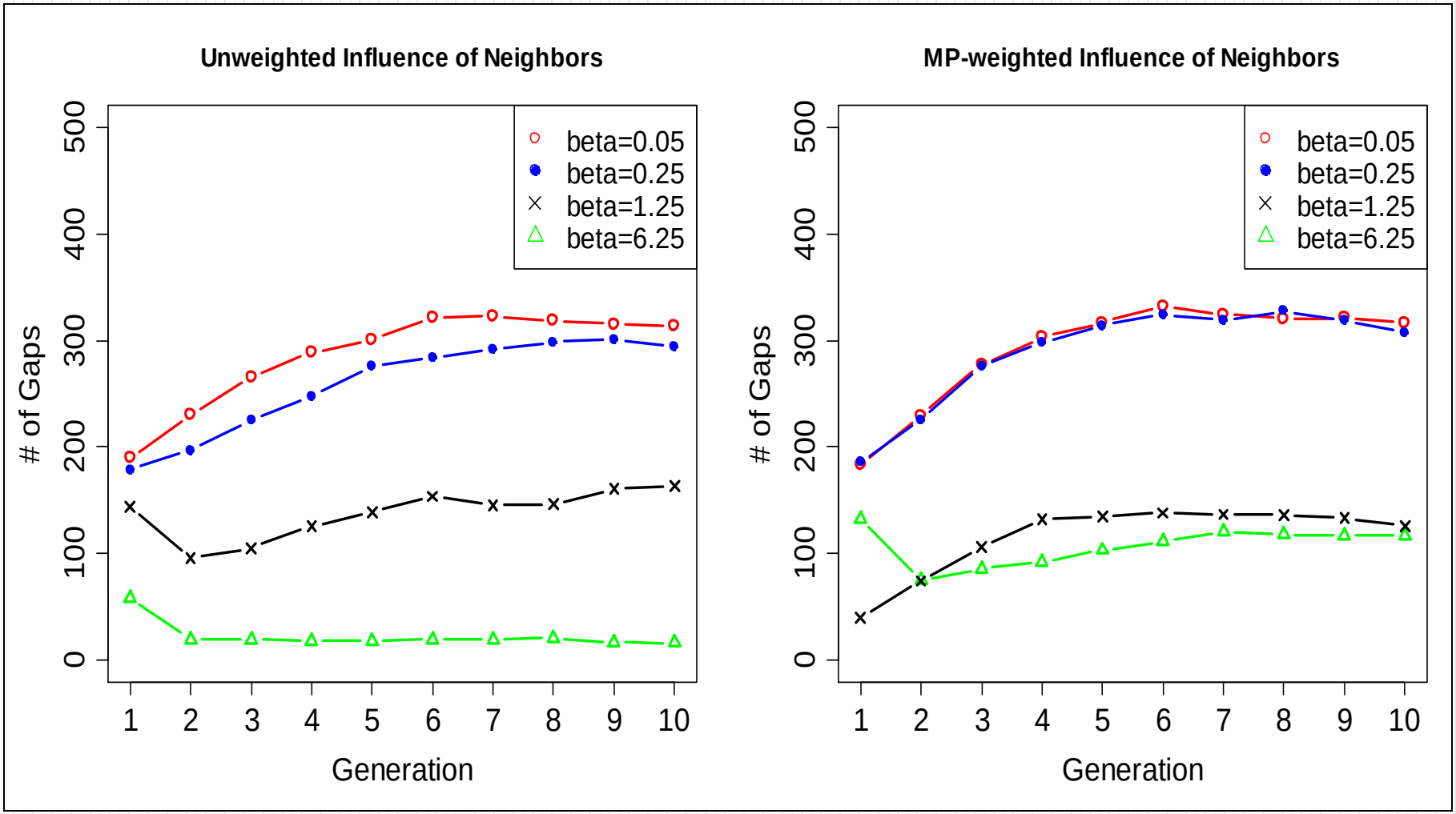
3) Observed relative freq

Raw tokens (ubedit')	1	53	210	27	71	91
Observed relative freq (ubedit')	0.2	11.7	46.4	6.0	15.7	20.1

4) Mix, weighting by raw number of observations

Predicted relative freq (ubedit')	**0.3**	**11.6**	**46.4**	**6.0**	**15.6**	**20.2**

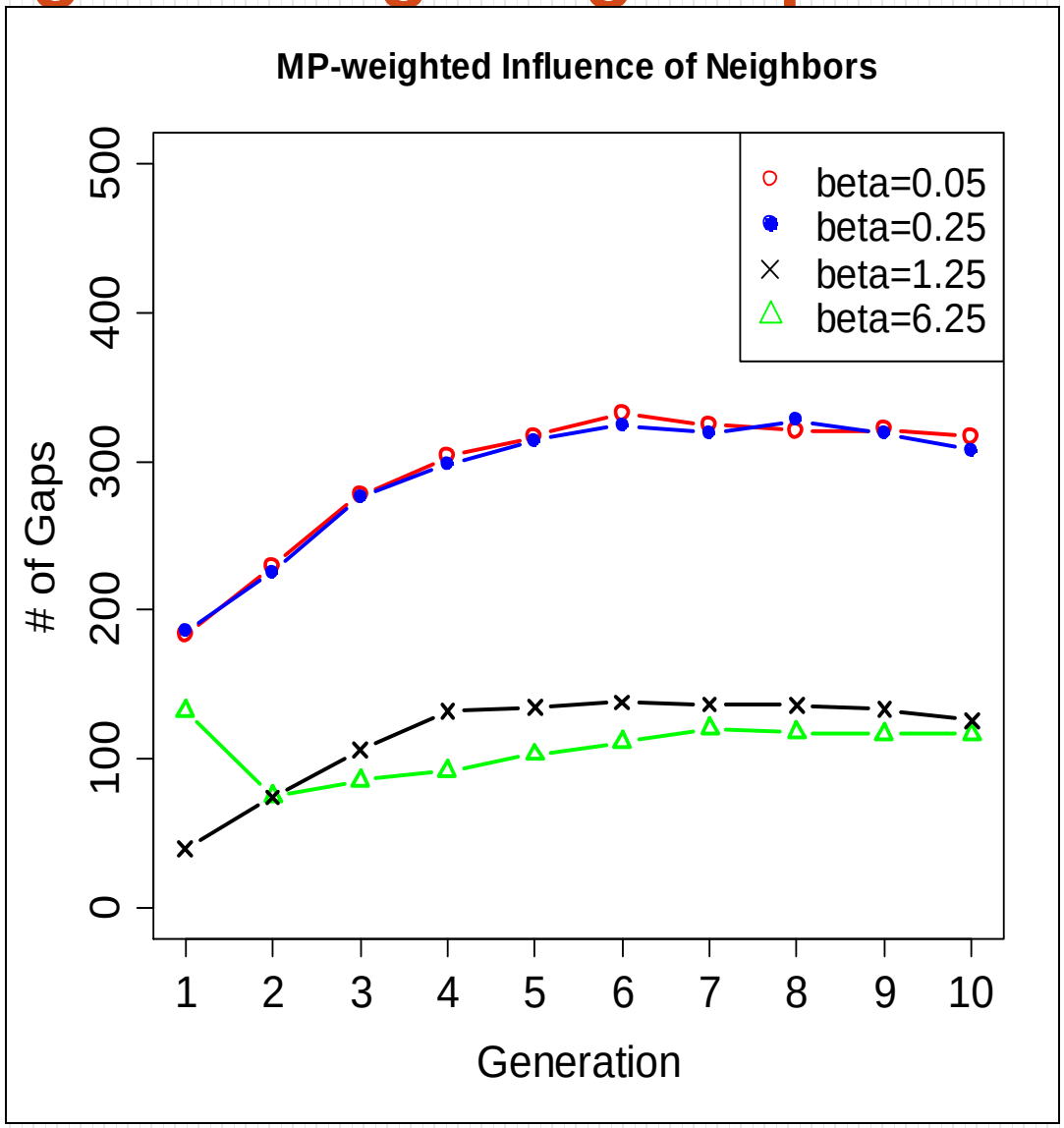
Results of simulation



Morpho-phonological weighting helps!

Weighting by morphophon. similarity increases the number of gaps in a generation when analogical influence is strong.

Number of gaps per generation increases, then reaches point of local stability.



Lesson to take away from simulations

- Successfully modeled the persistence of Russian paradigmatic gaps – but only when behavior of morphophonologically similar neighbors was given greater weight
- Conclusion: Not random that gaps follow distribution of alternation.
  - Morphosyntactic distribution (low 1sg relative frequency) promoted by morphophonological coherence
  - Defectiveness as the *expected* behavior
- This is fundamentally similar to a lexical gang effect
  - Compare to mild productivity of some English irregular past tense gangs

3. Sensitive to structure of the lexicon?

Some concluding thoughts

- Paradigmatic gaps are remarkably similar in some respects to more 'normal' inflectional patterns.
- Perhaps this shouldn't surprise us; 'canonical' defectiveness involves lexical specification, and many classically morphological traits are tied to the organization of the lexicon
  - See recent line of work suggesting that all of the oddballs of the inflectional system (e.g. suppletion) aren't so odd at all. Inflectional defectiveness is, probably, the most odd of them all – an outright failure of inflection. But it may just be the end of the cline.
  - Manifests morphosyntactically, rather than morphophonologically
- But why defectiveness? If defectiveness is so similar to other kinds of inflectional phenomena, then why do we have defectiveness at all? It is disruptive to the linguistic system...
- Keeping a clear eye on the speaker – I speculate that the difference between analogical extension and defectiveness rests in social conditions, more than in the structural preconditions

Thank you!

4. Prone to lexicalization?

What we are looking for:

covert reanalysis of defectiveness as a lexical property

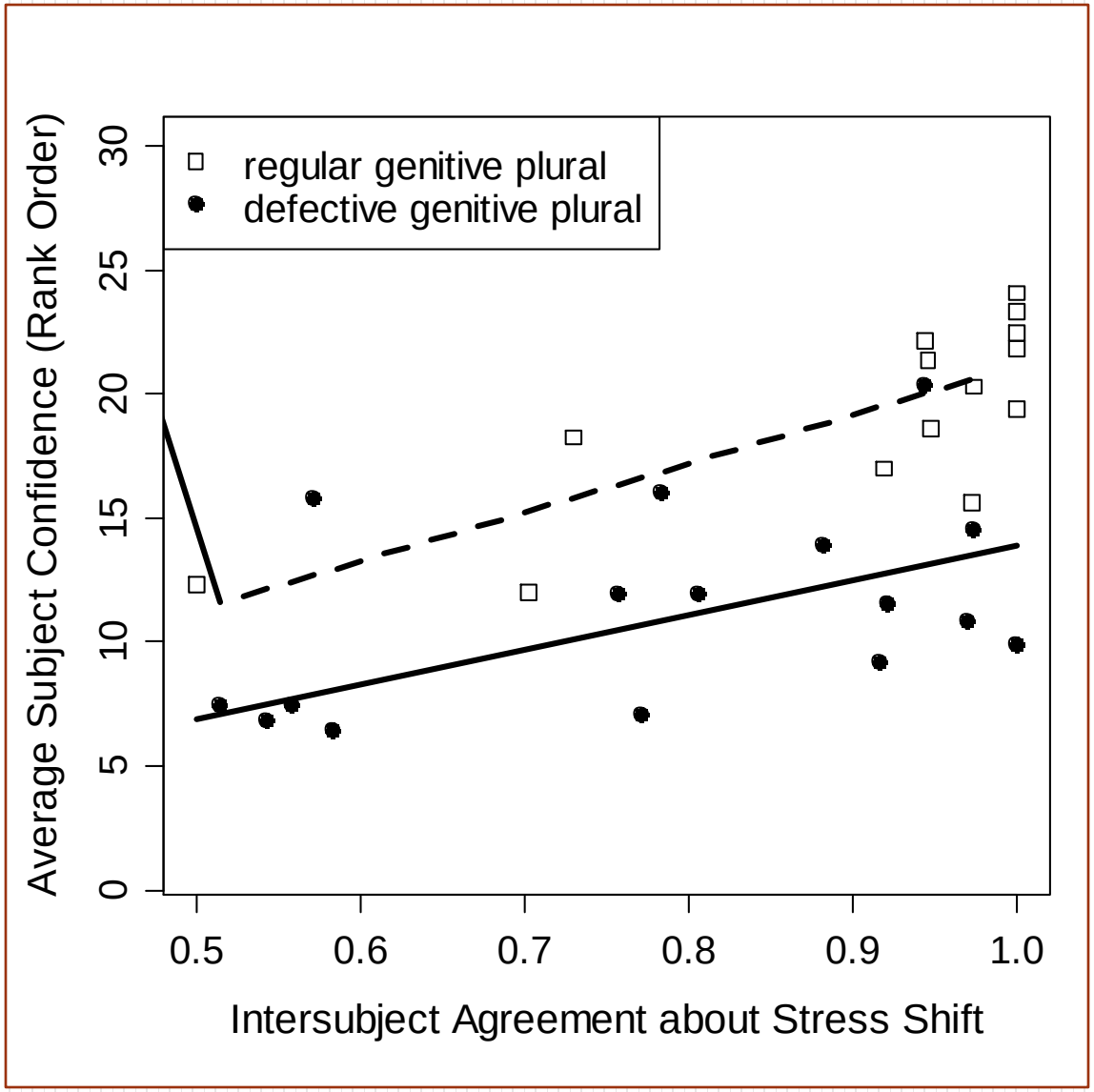
Greek gaps: Experimental methodology

- Replication of Albright (2003)'s methodology for studying Spanish verbal gaps
- Experiment with three tasks
  - Word familiarity judgment
  - Cloze procedure – produce word-form for sentence context
  - Self-rating of confidence in production in cloze procedure
    - (Percentage scale, later converted to ranking of items)
- Albright's primary conclusion: Two patterns of Spanish verbal gaps fall out epiphenomenally from competition among inflectional patterns and frequency sensitivity
  - Correlation between interspeaker agreement and self-rating of confidence in production
  - Single continuum for defective and non-defective lexemes
    - Defective → least interspeaker agreement and lowest confidence

Greek gaps: Experimental methodology

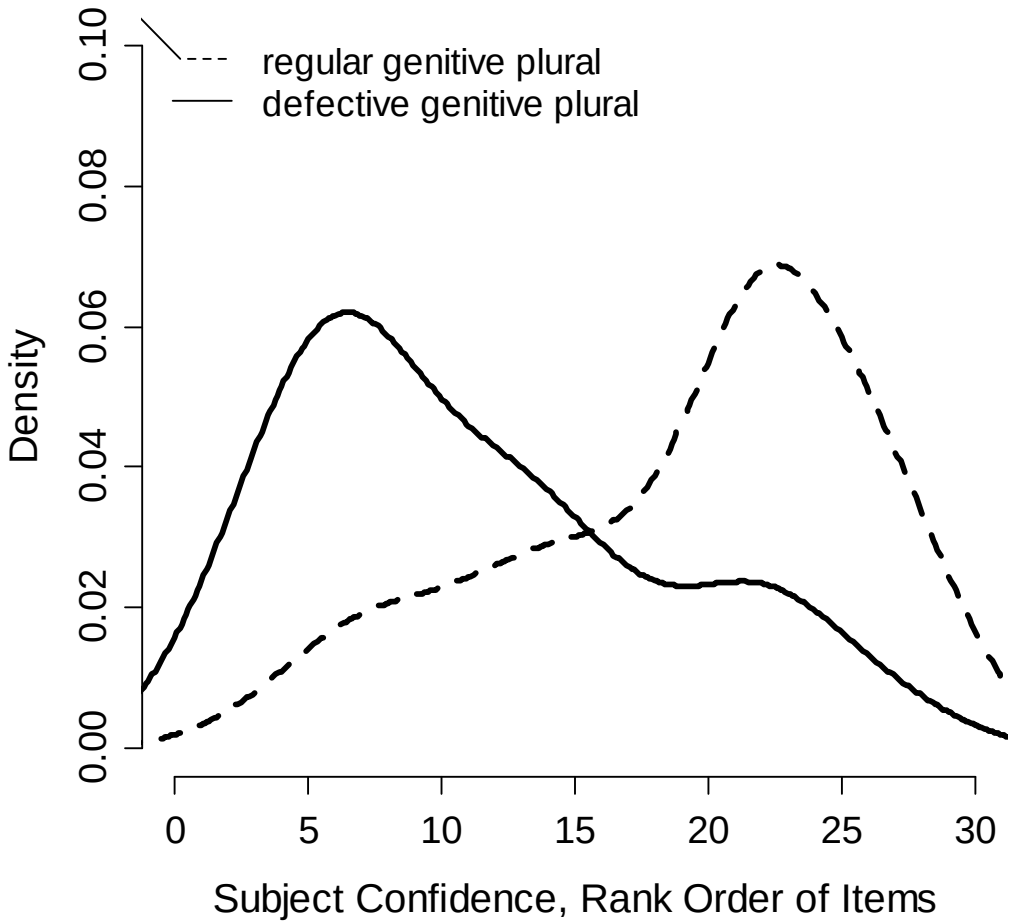
- Two forms elicited for each target lexeme
  - Genitive plural
  - Nominative singular or nominative plural
- 30 target lexemes, split between normatively defective and normatively non-defective
  - Comparable frequencies in Hellenic National Corpus
  - All belonged to class in Table 1
- 35 native Greek speakers
  - All students at Aristotle University of Thessaloniki
  - 31 females, 4 males

Prone to lexicalization



Correlation between intersubject agreement and confidence self-ratings. However, no single continuum for defective and non-defective lexemes.

Prone to lexicalization



Defective items receive lower ratings than do non-defective items

Conclusion: Covert reanalysis of Modern Greek gaps as lexical generalizations

Contra Albright's conclusions for Spanish

4. Prone to lexicalization?