Schema extraction in usage-based morphophonology

Vsevolod (Volya) Kapatsinski

University of Oregon

http://pages.uoregon.edu/vkapatsi/

Usage-based morphophonology

- Usage-based
  - Empiricist: Try to explain the emergence of phonological patterns on the basis of experience with language (Bybee 2001).
  - Central question:
    - Given a certain experience, what generalizations / representations does the learner update? In what direction? By how much?
    - Conversely, what experiences lead to the formation of the kinds of generalizations we see being productively used in human languages?
      - This is what miniature artificial language learning is good for (not for testing learnability)

- *Morph*ophonology
  - The part of phonology that interfaces with meaning / can be captured by constructions (form-meaning pairings) and mappings between constructions

Schema (Nesset 2008)

- First-order schema: A recurrent form-meaning pairing.
  - E.g., "plurals end in tʃi#"
  - Proposed to be the basic unit of grammar, including morphology (Bybee 1985, 2001, Bybee & Moder 1983, Bybee & Slobin 1982, Nesset 2008)
  - =Construction (Booij 2010, Goldberg 1995, 2002, 2006)
  - =Product-oriented schema (Bybee 1985, 2001; Bybee & Slobin 1982)
  - Focus of this talk
- Second-order schema: A mapping between first-order schemas
  - E.g., SG=…k# ~ PL=…tʃi#
  - Captures arbitrary paradigmatic mappings
    - We need them to describe productive patterns in natural languages: Booij (2008, 2010), Gouskova & Becker (2013), Nesset (2008), Pierrehumbert (2006)
    - But they are really hard to acquire and seem to require prior acquisition of first-order schemas: Frigo & MacDonald (1998), Weinert (2009), &c
  - =allostructions, Cappelle (2006)

Schemas are not all there is

- Central feature of language to account for by any theory of grammar: Creativity (Chomsky 1959)
- For morphology, what happens when a speaker derives a novel form of a word from a known form of the same word?
  - Paradigm Cell Filling Problem (Ackerman et al. 2009)
- Point of present paper:
  - Part of what happens is the speaker *perseverates on* gestures and larger chunks comprising the known form
    - Could be captured with output-output faithfulness constraints (Benua 1997, Kenstowicz 1996, Nesset 2008) but understood as a production-internal phenomenon
    - Most of the time this increases accuracy: *most* of the form needs to be repeated (Ambridge & Lieven, 2011; Farrar, 1992; Rubino & Pine, 1998)
      - cf. Rumelhart & McClelland 1986: *mail-membled* vs. Pinker & Prince (1988)
  - Given perseveration, schemas should start out general and become more specific with learning (*a la* McClelland et al., 1995)

Evidence for first-order schemas

Kapatsinski (2012, 2013)

Rules: Changes in context

- Consider these rules for making plurals:
  - k→tʃi
  - t→ti
  - p→pi
  - tʃ→tʃi

- Can you come up with more general rules?
  - 0→i/__#
  - k→tʃ/__i

- What rule do examples like mutʃ→mutʃi support?
  - 0→i/__#
  - And maybe 'do nothing /___i' (Albright & Hayes 2003)

- Palatalization should be hurt, or at least not helped by these examples

- Assumptions (Albright & Hayes, 2003):
  - Generalizations made over *word pairs*
  - *Decomposed into change and context*
    - Form comparison necessary (cf. Mitroff et al. 2004 "nothing compares two views")
    - Different from second-order schemas

Prince & Smolensky (1993)

Optimality Theory

- Consider these rules for making plurals:
  - k→tʃi
  - t→ti
  - p→pi
  - tʃ→tʃi

- How would we describe this pattern in OT?
  - Markedness constraint: *ki (or *k or *Stop-i or *VStopV)
  - Faithfulness constraints: Ident-[Dorsal], Ident-[Labial], Ident-[+anterior], Ident-[DelRel]
  - *ki >> Ident-[Dorsal], Ident-[DelRel]

- What ranking do examples like mutʃ→mutʃi support?
  - They say nothing about *ki
  - If anything, they support Identity constraints

  - Palatalization should be hurt, or at least not helped by these examples

Why am I asking?

- Because examples like mutʃ→mutʃi are actually taken by learners to support the generalization that plurals should end in [tʃi]…

The artificial languages

	tapa	tipi
k→tʃi	100% 4	
{t;p} → {t;p}i	25% 2	75% 6
{t;p} → {t;p}a	75% 6	25% 2

Two plural suffixes –i and -a
If –i attached to a **velar** ({k;g}), the **velar** changes to an **alveopalatal**
This is **velar palatalization**

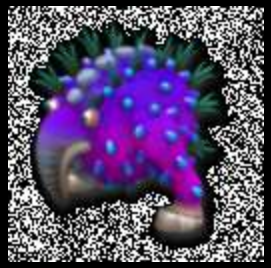
The artificial languages

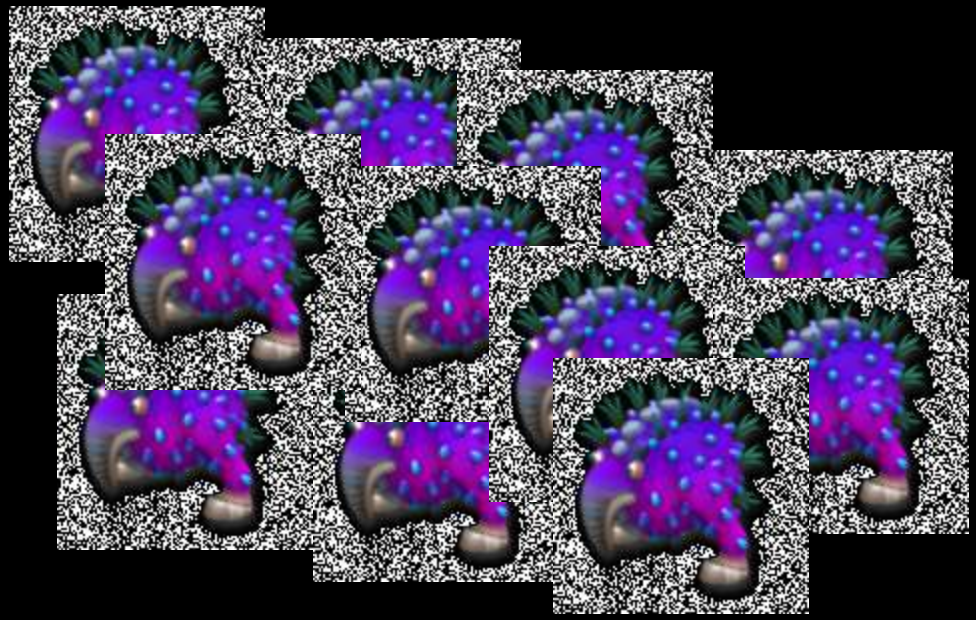
	tapa**chi**	tipi**chi**
k→tʃi	100% 4	
tʃ→tʃi	100% 4	
{t;p} → {t;p}i	25% 2	75% 6
{t;p} → {t;p}a	75% 6	25% 2

Task: Repeat singular-plural pairs

Goal: making singular-plural relationships obvious to facilitate rule extraction

Video:						
Audio:		[boʊk]			[boʊtʃi]	
Learner action:	Watch	Watch & listen		Watch	Watch & listen	Repeat aloud, then click
Duration:	300 ms	500-900 ms	500 ms	300 ms	500-900 ms	0-10s





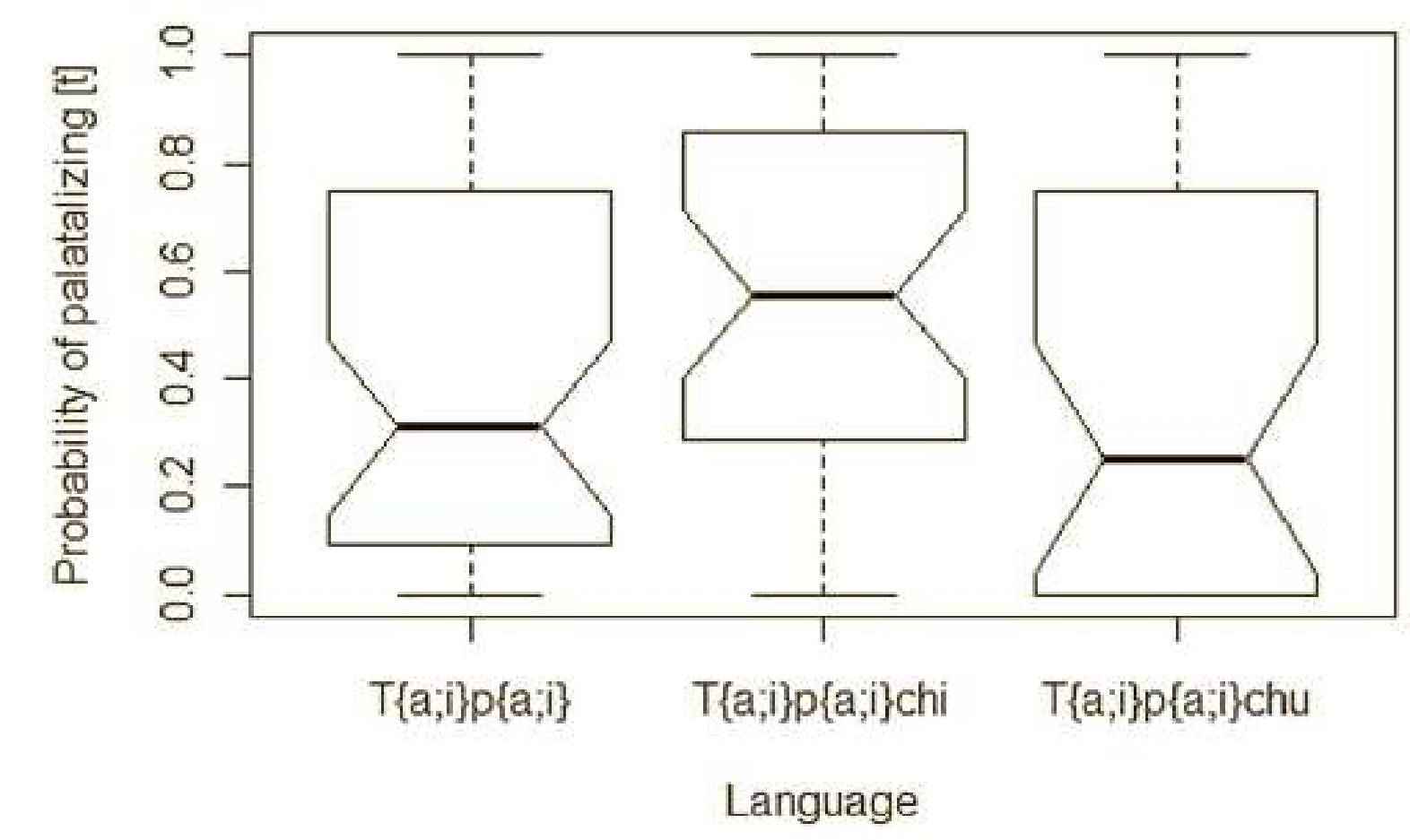
Elicited production

Video:				
Audio:		[vik]		
Learner action:	Watch	Watch & listen		Say the plural aloud, then click
Duration:	300 ms	500-900 ms	500 ms	0-10s

Berko (1958)

Support for first-order schemas

t→tʃi is favored by examples of tʃ→tʃi



Kapatsinski (2013)

So… what mapping is tʃ→tʃi most like?

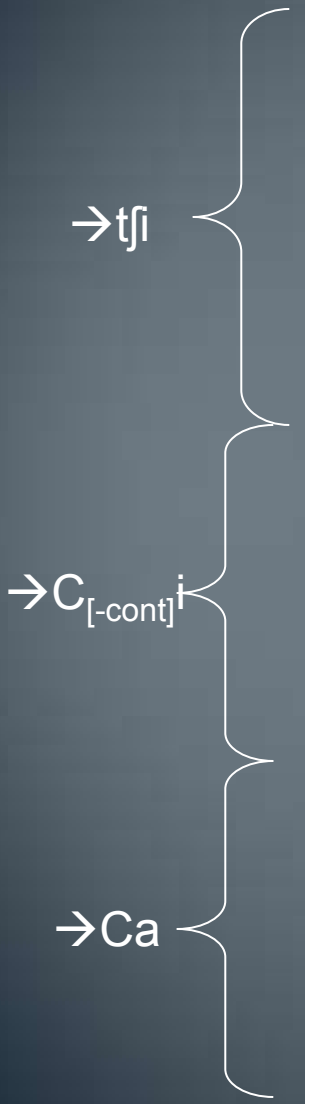
- Rules: t→ti, p→pi, k→ki (all 0→i/C_)
- Schemas: k→tʃi, t→tʃi, p→tʃi (all → tʃi)
- How to test:
  - If use of two mappings is due to the same underlying generalization, the productivities of the mappings must covary across subjects
    - Those who weight the generalization high, should apply both mappings a lot
    - Those who weight the generalization low, should not apply both mappings

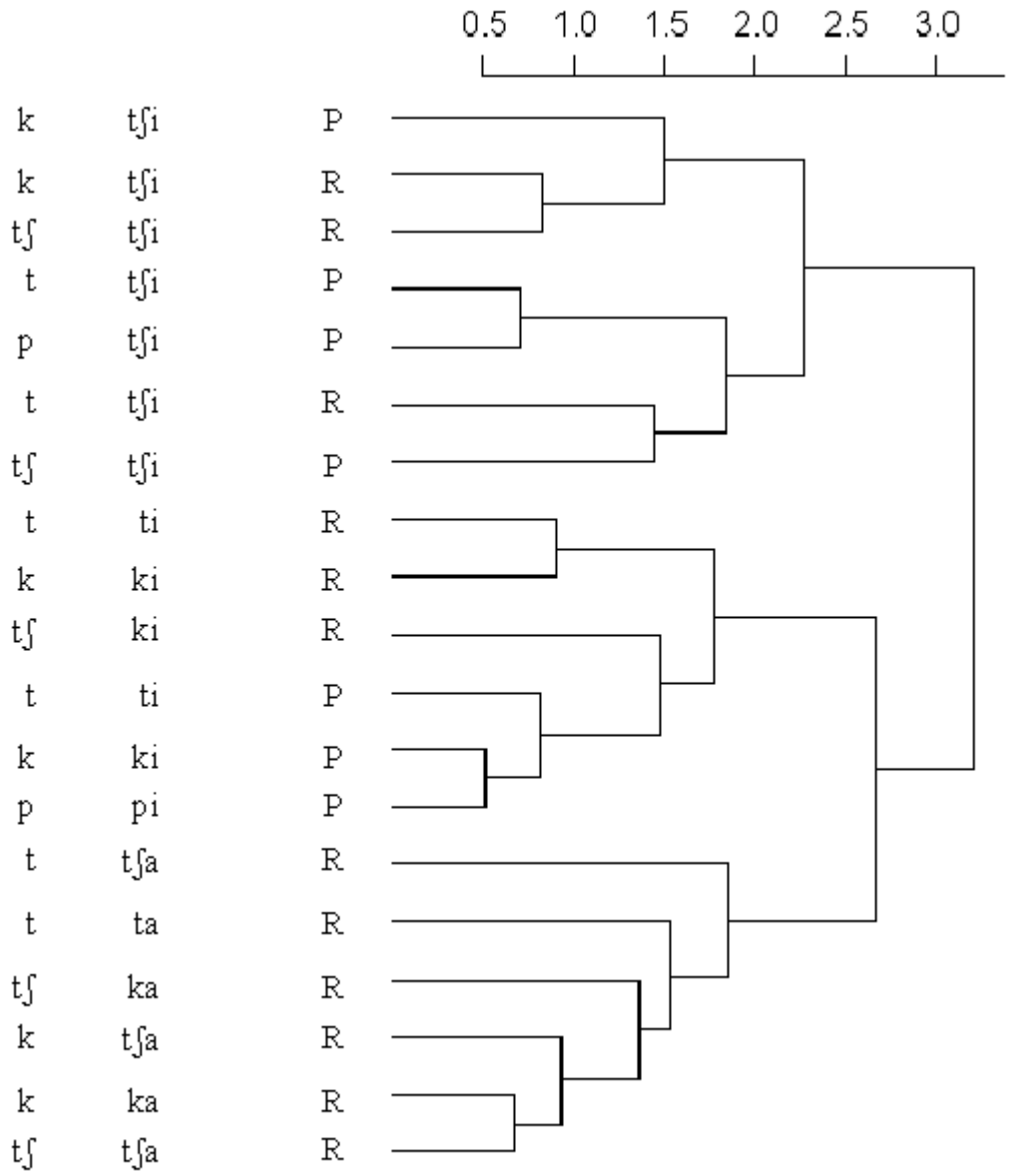
	Subject1	Subject2	Subject3	Subject4	Subject5	Subject6
k→tʃi	10	1	20	5	8	3
k→ki	10	15	0	9	6	17
k→ka	0	5	0	6	5	0
k→tʃa	0	0	0	0	1	0
tʃ→tʃi	15	0	20	8	12	5
tʃ→tʃa	5	20	0	12	8	15
t→tʃi	8	0	10	1	2	0
t→ti	10	18	10	17	16	20

We can think of each subject as an independent dimension

Then we cluster the mappings on how similar they are on these dimensions

If subjects treat two mappings similarly (both either high or both low within a subject), they will be similar and clustered together





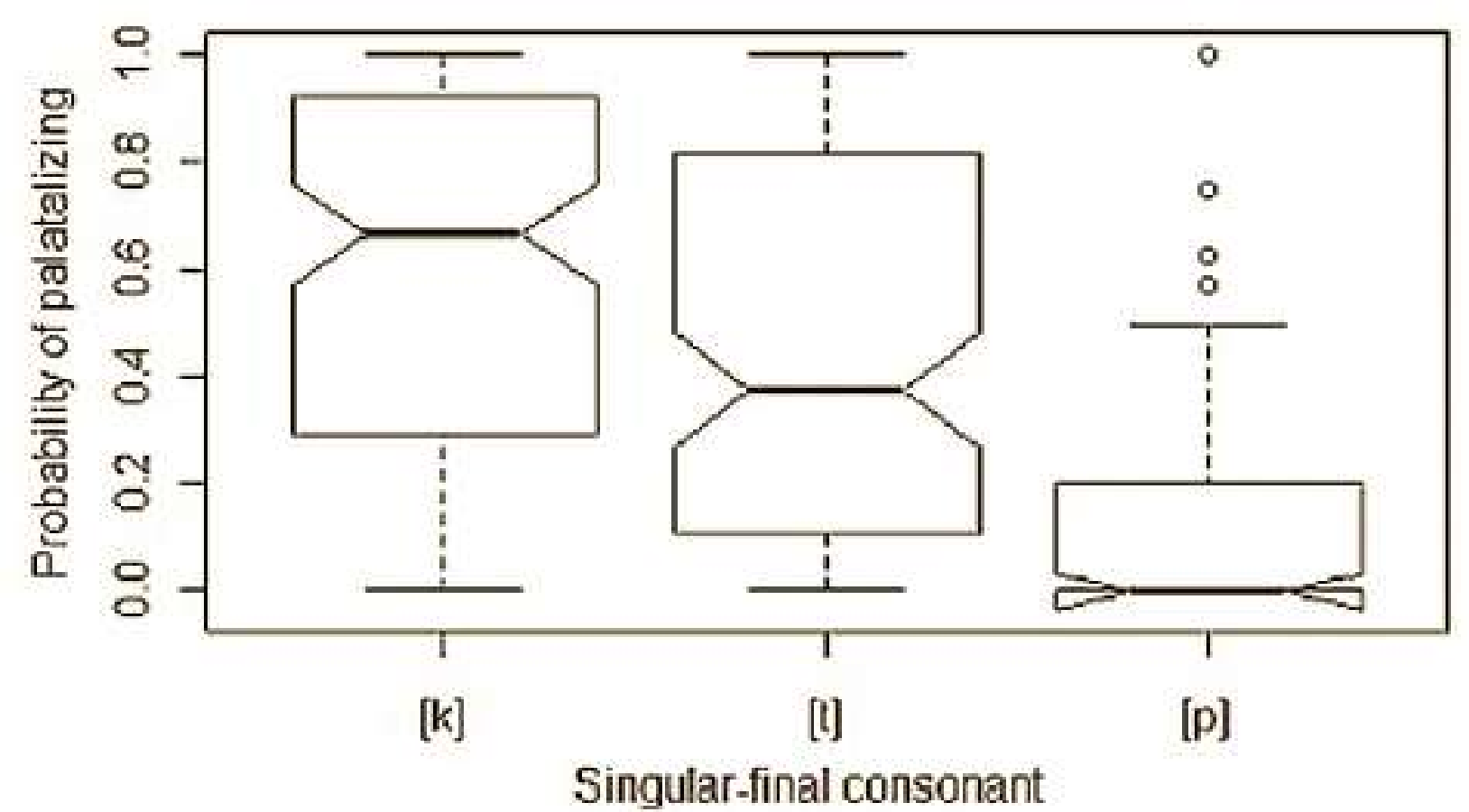
Kapatsinski (2012)

Interim conclusion

- Similarity in product is more important than similarity in mapping
  - Support for first-order schemas
- Even in a training paradigm that makes reliance on rules maximally easy
- Problematic for models that categorize paradigmatic patterns based on changes they feature
  - Rule-Based Phonology (Albright & Hayes 2003, Plag 2003, Reiss 2008; cf. Kisseberth 1970, Bybee & Slobin 1982, Prince & Smolensky 1993) and, in this case, OT (Prince & Smolensky 1993)

But… the learners *do* know *something* about the changes

Don't change (needlessly)



Kapatsinski (2013)

Inductive bias against labial palatalization

Labial:	Velar:	Alveolar:
ip→itʃa	ik→itʃa	it→itʃa
ap→{api;atʃa}	ak→{aki;atʃa}	at→{ati;atʃa}
k→ka	t→ta	k→ka
t→ta	p→pa	p→pa

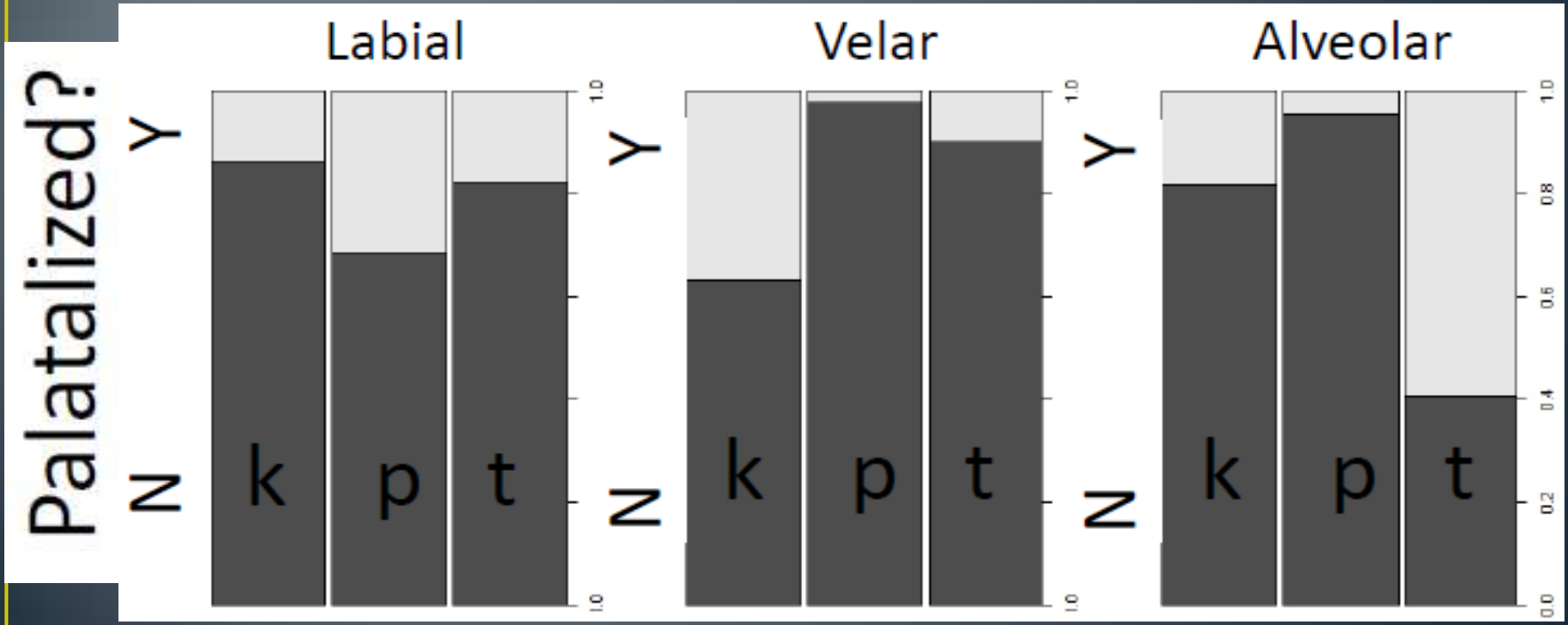
[p] has a labial gesture that is not preserved in [tʃ]
[tʃ] has both tongue blade ([t]) and tongue body ([k]) gestures (Yun 2006)

Stave, Smolek & Kapatsinski (2013)

Production data

Labial palatalization is not learned as well as the others in production

- Subjects exposed to labial palatalization tend to palatalize everything or palatalize nothing.
- No significant difference between the other conditions.



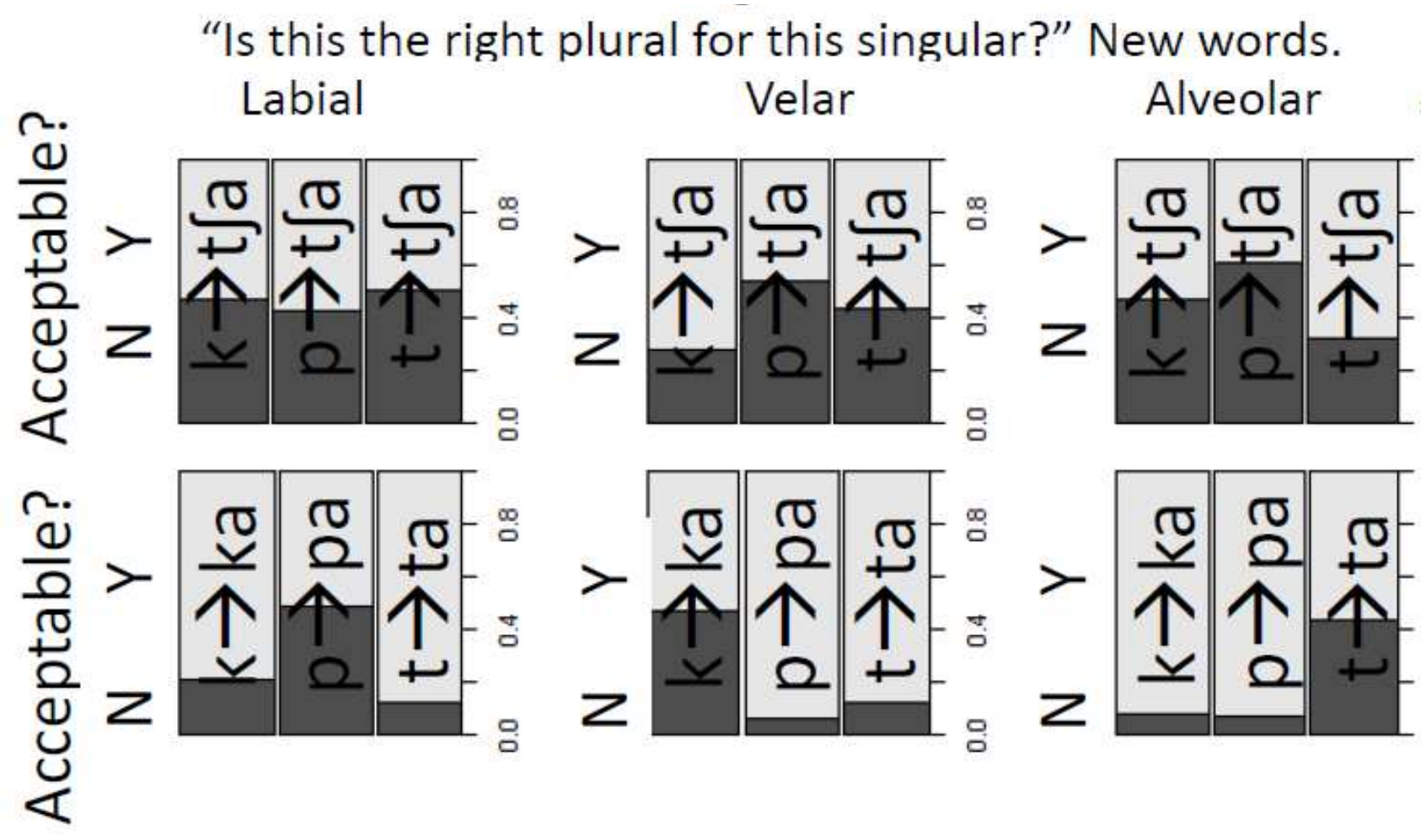
Stave, Smolek & Kapatsinski (2013)

Judgment data

Labial palatalization is not learned as well as the others in judgment

But differences are significantly smaller than in production

Consistent with the bias being fundamentally production-based

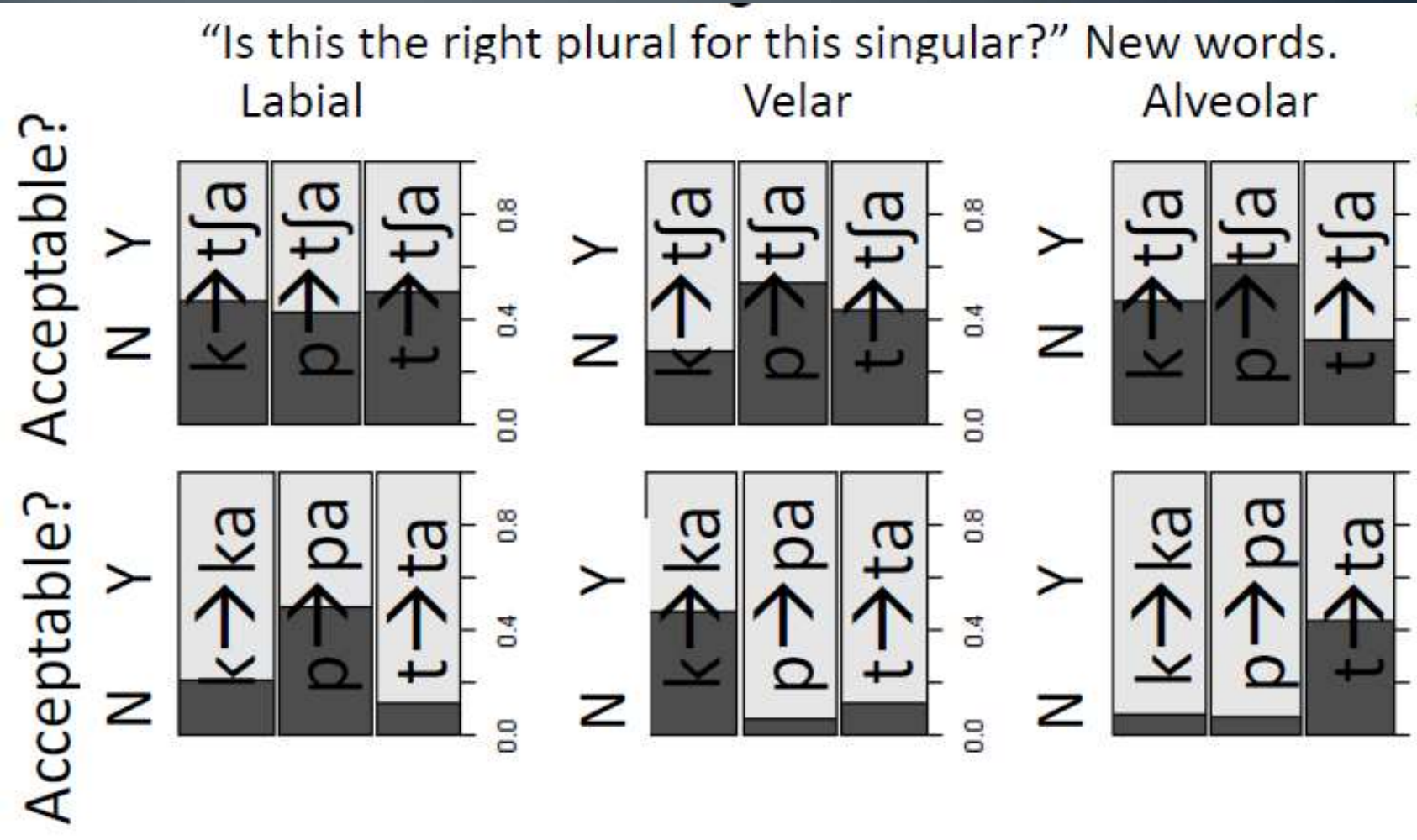


Stave, Smolek & Kapatsinski (2013)

Judgment data

The differences between conditions are about *changes*, not phonotactics:

Subjects can learn *ka, *pa and *ta equally well



Stave, Smolek & Kapatsinski (2013)

How I think it all works: Clamoring for Blending

How I think ~~it all~~ *some of it* works: Clamoring for Blending

Kapatsinski (2013), *Lg* 89(1)

Chunk-schema blending

- You get a singular form like [bup]
- Chunks from it clamor: “express me”
  - Chunk = gestures and up
- Some have a stronger voice:
  - Keep [p] >> Keep [t], Keep [k]
- Plural schemas: “express me”
- Schemas strengthen, and grow more specific during learning as more plurals are encountered
- Some have a stronger voice: PL=…tʃi#, PL=…[-DelRel]i#
  - PL=…tʃi# gets stronger as you hear a lot of {k;t;p;tʃ}→tʃi
    - mutʃi-mutʃi helps tuk-tutʃi and tut-tutʃi
- Stem changes are avoided until demanded by a strong, and specific, schema
  - PL=…tʃi# overtakes “Keep [t]”, “Keep [k]” before overtaking “Keep [p]”
- Telling error: bup→buptʃi: both “Keep [p]” and PL=…tʃi# win; PL=…Vtʃi# is not (yet) strong enough

Other support for the chunks clamoring for production

- "Addition bias" in speech errors: errors resulting in consonant clusters are more common than errors removing a consonant from a cluster (Stemberger 1991, Hartsuiker 2002)
- "Intrusion bias" in speech errors: the most common error if you look at articulator movements in tongue twisters like 'top cop' is simultaneous articulation of both stops (Goldstein et al. 2007)

Other support for the chunks clamoring for production

- "Preservation Principle": Almost categorical tendency to repair L1 phonotactic violations in L2 by insertion and not deletion (Paradis & LaCharité 1997, Kang 2011) unless overridden by strong schemas
  - Kang (2011): "all languages... that choose deletion repair in coda position have a strong preference for monosyllabic morphemes. But... even these languages do not systematically prefer deletion for onset clusters."

Evidence for schemas growing more specific over the timecourse of learning

General→Specific or Specific→General?

- Do we learn what particular segment sequences are possible and generalize? (standard view in C(x)G, e.g., Langacker 1987, Goldberg 2006, Tomasello 2003)
  - =start out with nothing being a legal plural…
- Or do we start out thinking anything goes and then narrow down on the classes of sequences allowed (=sequences observed)? (McClelland et al. 1995, Rogers & McClelland 2004: Progressive Differentiation)
  - For kids and adults learning a new generalization but not a new phoneme (sequence): everything is accepted at first but no schemas are strong enough yet to overcome perseveration
  - Kids show more perseveration, i.e., higher weights on "Keep X" → stem changes are disfavored in production (compared to adults) but not in judgment (Stave, Smolek, Kapatsinski & Redford, in progress)

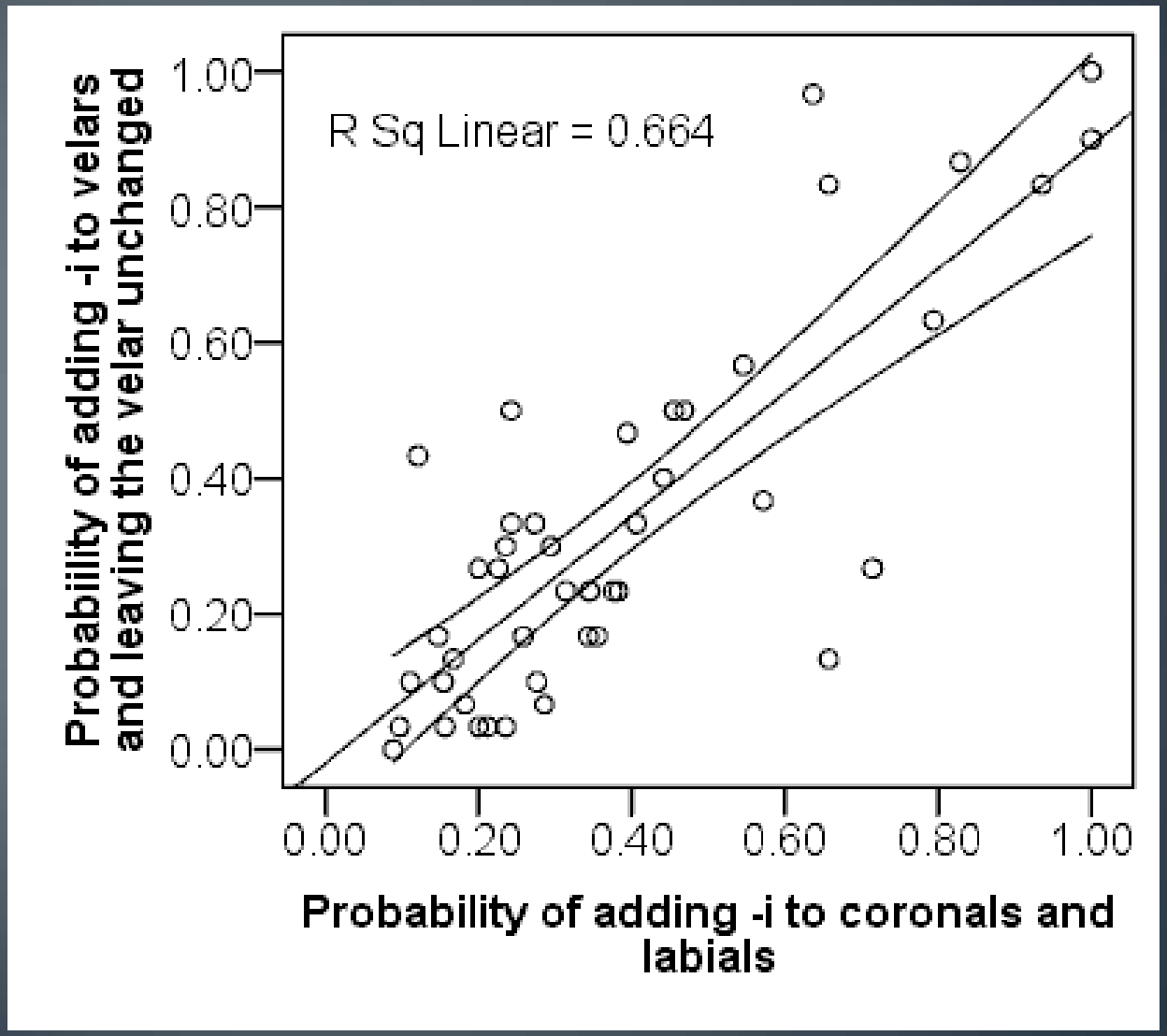
Emergence of changes and Automatic overgeneralization to natural classes

- Perseveration + ???=PL No change
- Perseveration + …i#=PL p→pi
- Perseveration + …tʃi#=PL p→ptʃi
- Perseveration + …Vtʃi#=PL p→tʃi

- [ti] and [pi] help [ki]

- see also Goldrick (2004), for phonotactic learning in production; DeJong et al. (2009) and Finley & Badecker (2009) for perceptual learning; Xu & Tenenbaum (2007) for word meaning learning; Ambridge et al. (2012) for syntactic constructions

Add –i to {t;p} → Add –i to k



Kapatsinski (2010a, 2010b)

Advantage of General→Specific: Automatic overgeneralization to natural classes

- [ti] and [pi] help [ki]
- Why not learn PL=…ti# & PL=…pi#?
  - {t;p} is not a natural class
  - Start with 'anything goes', look for unexpectedly frequent subclasses
    - CV → [-Del.Rel;-cont]i → [pi], [ti], [tʃi]

More evidence for general→specific learning

- Assuming all schemas are like words: when you don't know a word well, you are not sure about its form, accept major feature-changing mispronunciations as being the same word (Charles-Luce & Luce 1990, Pater et al. 2004, Shvachkin 1948/1973, Stager & Werker 1997, Swingley 2007, Swingley & Aslin 2007)

Implementation

A distributed connectionist network may work too (and maybe even better)

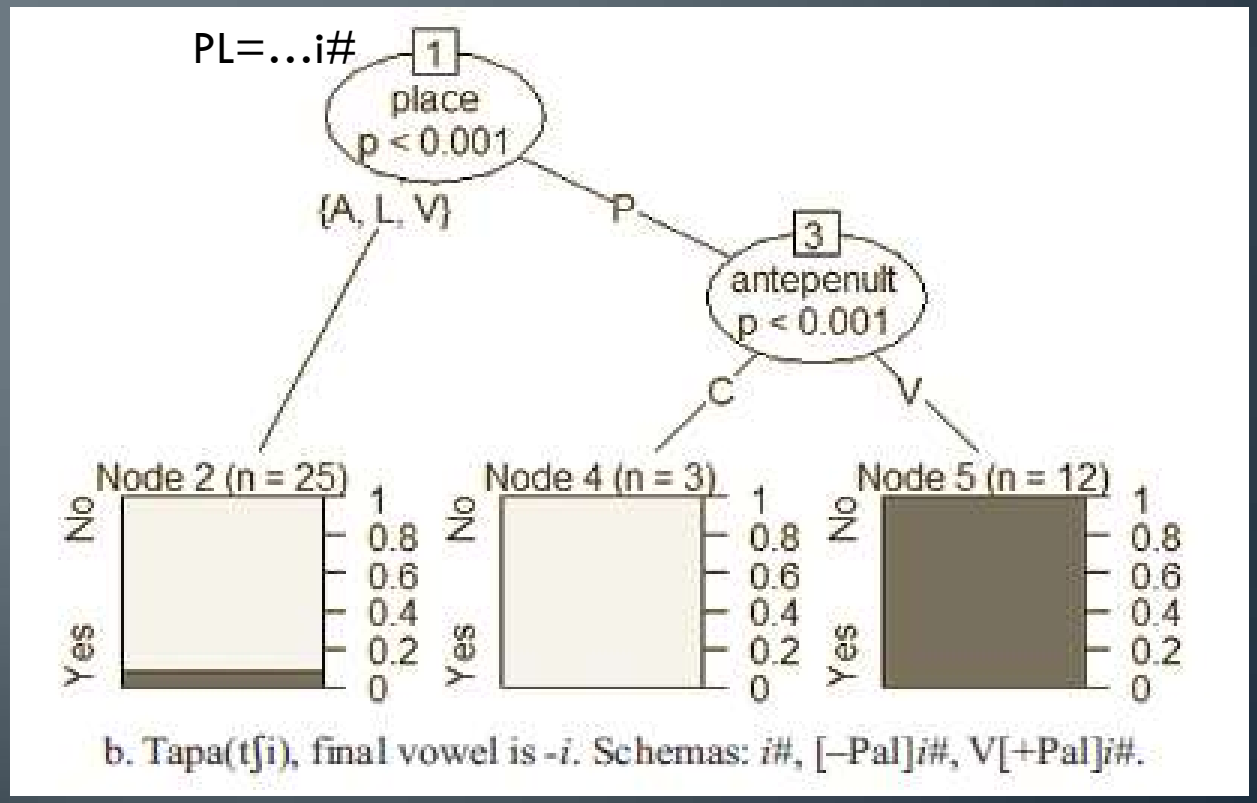
Conditional inference trees

- Predictors: phonological features of a wordform-final trigram
- Predicted: Is trigram observed in plurals?
  - Schemas as descriptions of forms with a certain meaning (Bybee 2001)
  - *Not* cues for predicting if form is singular or plural: adding tʃ→ tʃi makes tʃ less distinctive of plurality, yet favors it's production in plurals
- Algorithm: Recursive binary partitioning of the space of phonotactically legal wordforms (Hothorn et al. 2006) Using the most informative predictor feature at each step
  - Most informative: best distinguishes trigrams observed in plurals from others
  - General-to-specific
  - Good for small-n, large-p problems (as in a small lexicon): Strobl et al. (2009)
  - Expects non-cross-over interactions between features (all over the place in language: X has meaning Y IFF it has F_1 AND F_2 AND F_3…)
  - Bad at capturing cross-over interactions between features that are not of the same constituent (X has meaning Y IFF it has F_1 XOR F_2, which are hard to learn: Warker & Dell 2006, Kapatsinski 2009)
- ctree() in party package in R (Hothorn et al. 2006, Strobl et al. 2009)

Kapatsinski (2013)

Plural vowel is –i. Languages with many [ta] and [pa]

- A schema is a path that proceeds downwards from the root of the tree terminating in a node that is either 1) a leaf with a non-zero type frequency or 2) an ancestor to at least one leaf with a non-zero type frequency.

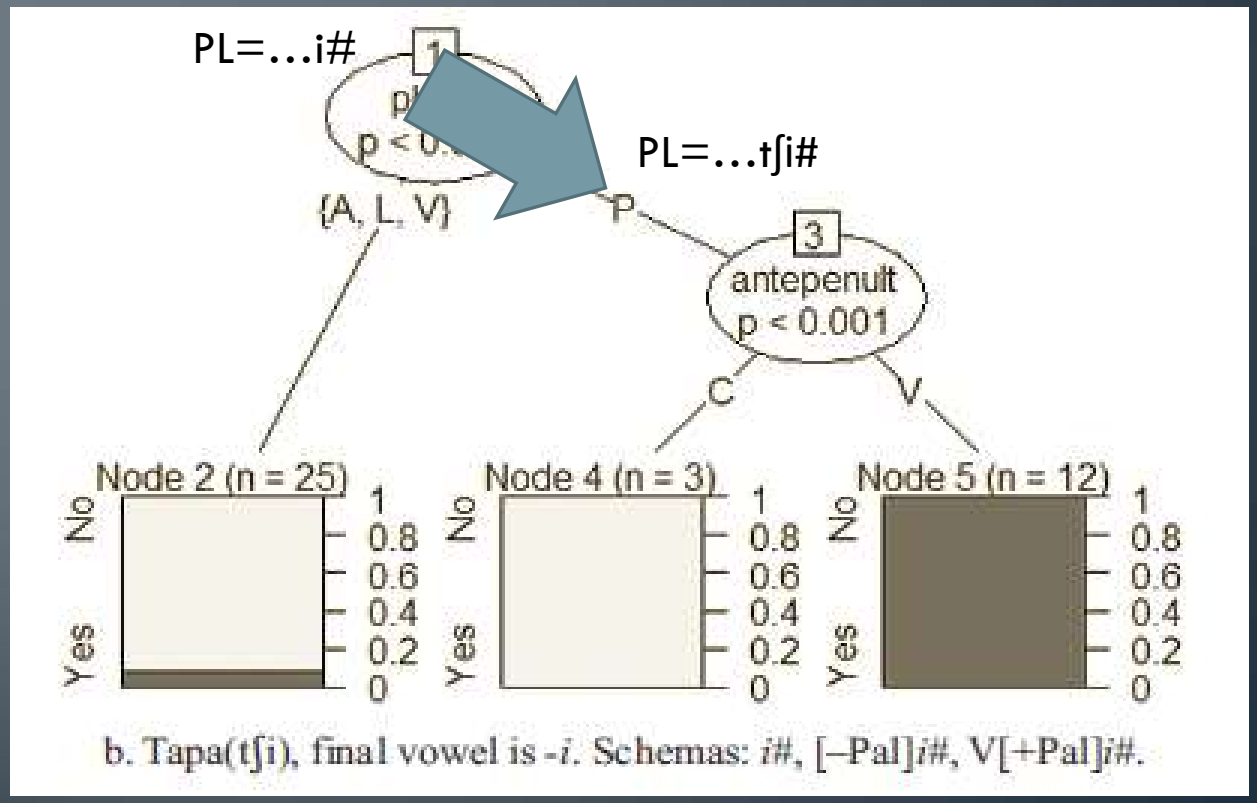


b. Tapa(tʃi), final vowel is -*i*. Schemas: *i*#, [–Pal]*i*#, V[+Pal]*i*#.

Kapatsinski (2013)

Plural vowel is –i. Languages with many [ta] and [pa]

- A schema is a path that proceeds downwards from the root of the tree terminating in a node that is either 1) a leaf with a non-zero type frequency or 2) an ancestor to at least one leaf with a non-zero type frequency.

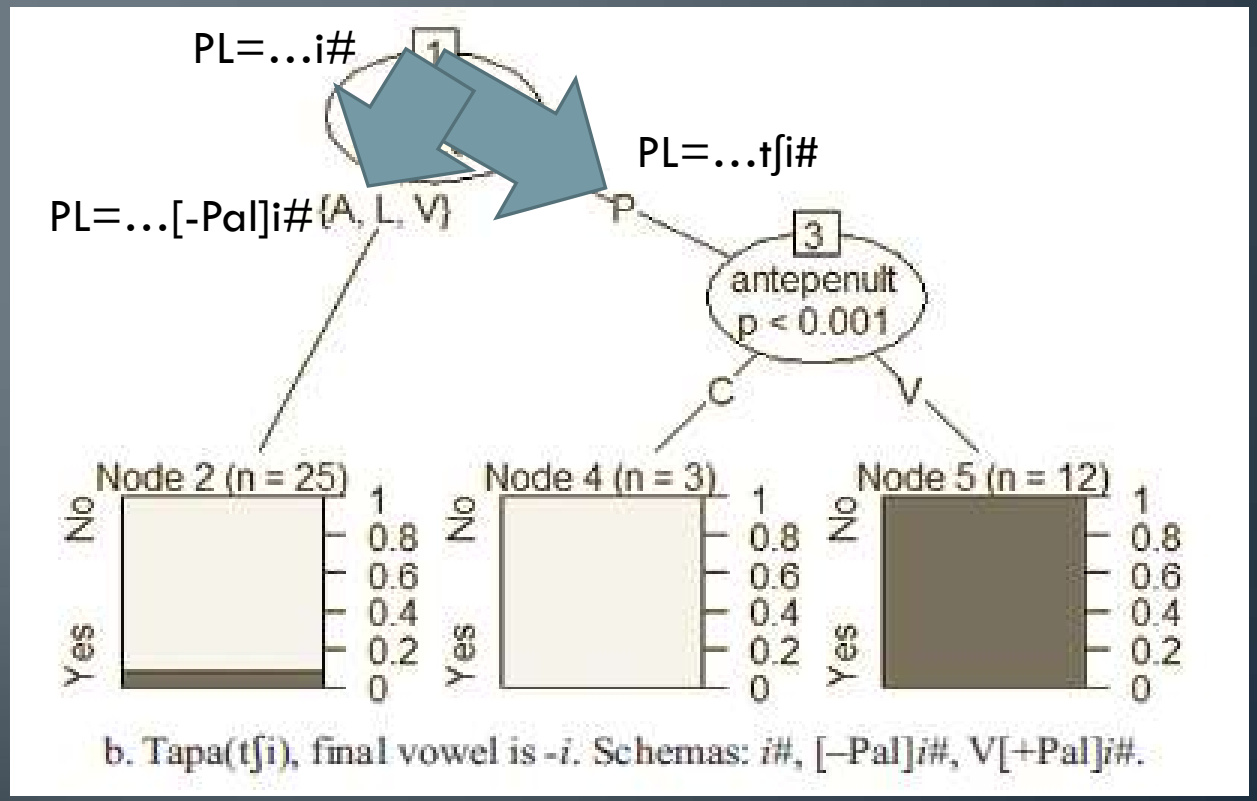


b. Tapa(tʃi), final vowel is -*i*. Schemas: *i*#, [–Pal]*i*#, V[+Pal]*i*#.

Kapatsinski (2013)

Plural vowel is –i. Languages with many [ta] and [pa]

- A schema is a path that proceeds downwards from the root of the tree terminating in a node that is either 1) a leaf with a non-zero type frequency or 2) an ancestor to at least one leaf with a non-zero type frequency.

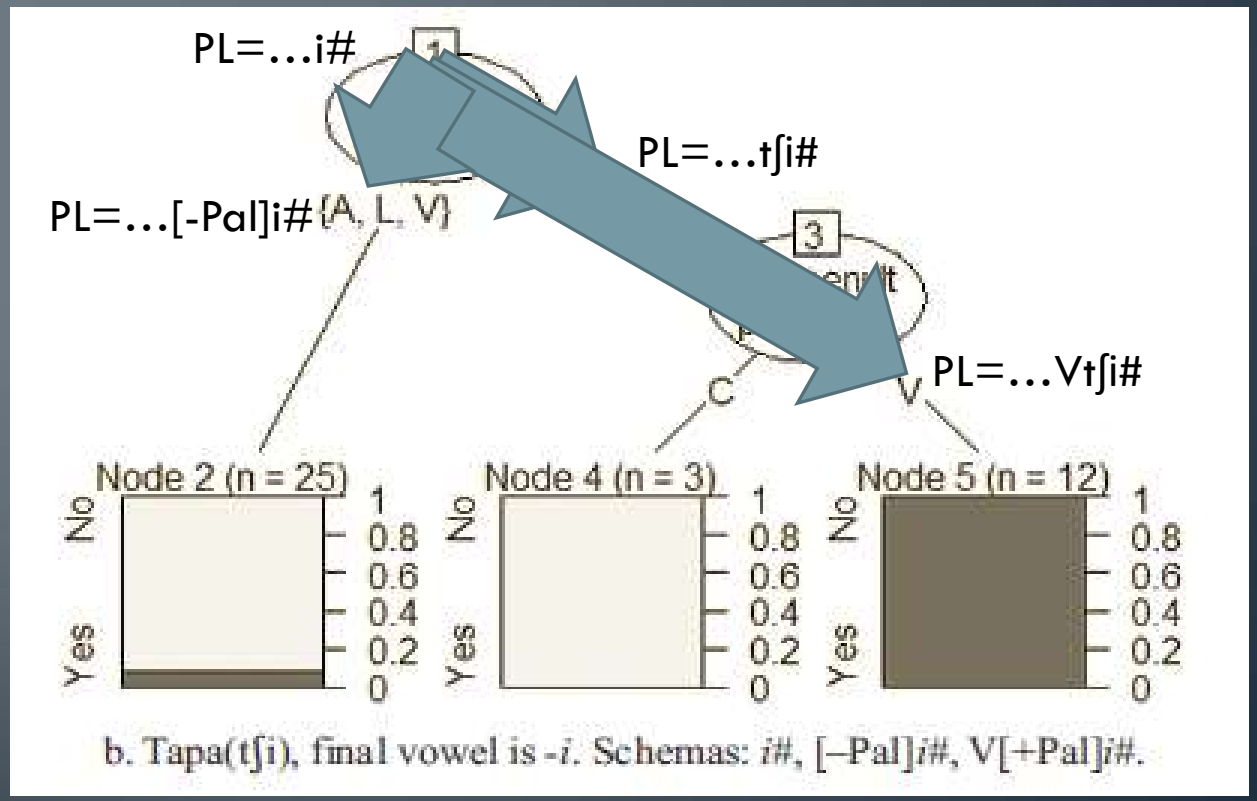


b. Tapa(tʃi), final vowel is -*i*. Schemas: *i*#, [–Pal]*i*#, V[+Pal]*i*#.

Kapatsinski (2013)

Plural vowel is –i. Languages with many [ta] and [pa]

- A schema is a path that proceeds downwards from the root of the tree terminating in a node that is either 1) a leaf with a non-zero type frequency or 2) an ancestor to at least one leaf with a non-zero type frequency.



b. Tapa(tʃi), final vowel is -*i*. Schemas: *i*#, [–Pal]*i*#, V[+Pal]*i*#.

Kapatsinski (2013)

Conclusions

- A speaker of a language presented with a new singular form can create a new plural form from it.
- How do speakers do this?
  - Clamoring for Blending (C4B): Competition between chunks perseverating from the known form and schemas defining forms with the intended meaning
  - Plus, second-order schemas, phonotactic constraints?
- Which generalizations will be supported by a given perceptual or production experience, given the learner's prior experience and inherent bias, and which other generalizations will lose strength as a result of that experience.
  - Experience a plural form → learn what plurals are like
  - If plurals have a consistent form, then eventually the pressure to produce plurals that are like other plurals in the language can override even strong perseveratory tendencies

Speculations

- What experiences lead to second-order schemas, constraints?
  - Experience a singular form and a plural form → learn which chunks should and shouldn't be retained in forming plurals (or singulars)
  - Experience difficulty producing a form → avoid that form and such gestures (Berg 1998, Martin 2007, Redford 2008, Schwarz & Leonard 1982)
  - Produce a novel form from a known form → learn second-order schemas?

Write-ups of parts

Kapatsinski, V. 2013. Conspiring to mean: Experimental and computational evidence for usage-based morphophonology. Forthcoming in *Language*, 89(1).

Kapatsinski, V. 2012. What statistics do learners track? Rules, constraints or schemas in (artificial) grammar learning. In Gries, S. Th., & D. Divjak, eds. *Frequency effects in language learning and processing*. Mouton de Gruyter.

Kapatsinski, V. 2010a. Velar palatalization in Russian and artificial grammar: Constraints on models of morphophonology. *Journal of Laboratory Phonology*, 1(2), 361-393.

Kapatsinski, V. 2010b. Rethinking rule reliability: Why an exceptionless rule can fail. *Chicago Linguistic Society*, 44(2), 277-291.

Kapatsinski, V. 2009. Testing theories of linguistic constituency using configural learning: The case of the English syllable. *Language*, 85(2), 248-77.

Stave, M., A. Smolek, & V. Kapatsinski. 2013. Inductive bias against stem changes as perseveration: Experimental evidence for an articulatory approach to output-output faithfulness. Submitted to *Proceedings of the Annual Conference of the Cognitive Science Society*.