Effects of Lexical Frequency and Phonotactics on the Learning of Phonological Alternations

Gaja Jarosz

Yale University

Learning Meets Acquisition Workshop
Osnabrück, Germany, March 2009

Outline

- Zamuner, Kerkhoff, & Fikkert (in prep):
  - Investigate interaction of phonotactics & alternations
  - What can explain the complex effects?
- Computational model
  - The Learning Theory (Jarosz 2006)
  - The simulation & explanation of effects
- Discussion

Phonotactics & Alternations

- There is reason to think learning of phonotactics may precede learning of alternations:
  - Experimental findings support this overall progression…
  - Phonotactic knowledge (including contrasts):
    - Some knowledge by **9 months** (Jusczyk et al 1993; Friederici and Wessels 1993; Anderson et al, 2003; Werker and Tees, 1984)
  - Alternations:
    - Up to **4 yrs** or even later before productive alternations are acquired (Berko 1958; Stager and Werker 1997; Pater 1997; Pater, Stager and Werker 2004; MacWhinney 1978; Fikkert & Freitas 2006; Bals, Odden & Rice 2005)
    - Though interesting new findings by White, Peperkamp, Kirk, & Morgan (2008)

Phonotactics & Alternations

- There is reason to think learning of phonotactics may precede learning of alternations:
  - Arguments from Learnability…
  - Much grammar learning can occur prior to knowledge of morphology and underlying forms (Tesar & Prince, 2003/2007; Prince & Tesar 2004; Hayes 2004, & others)
  - That knowledge can aid in later morphophonemic learning (Tesar and Prince, to appear; Jarosz 2006)
- But there is very little work addressing the interaction of phonotactics and alternations in acquisition…

Acquisition of Voicing Alternation

- Zamuner, Kerkhoff, & Fikkert (in prep: "ZKF") address the relationship between phonotactics and alternations directly
- Picture naming task focusing on [t] & [d] in Dutch nouns
- Two age groups: 2;6 and 3;6
- Production accuracy by target UR and morphological status:
  - Mono-morphemic target /t/
    - [wɑtər] "water"
  - Mono-morphemic target /d/
    - [rɪdər] "knight"
  - Bi-morphemic target /t/
    - [pɛtə] "caps" ~ [pɛt] "cap"
  - Bi-morphemic target /d/
    - [bɛdə] "beds" ~ [bɛt] "bed"

Acquisition of Voicing Alternation

- ZKF's results:
  - Children never produced [d] syllable-finally
    - /bɛd/ and /pɛt/ never pronounced with $d]_\sigma$

Acquisition of Voicing Alternation

- ZKF's results:
  - Children never produced [d] syllable-finally
    - /bɛd/ and /pɛt/ never pronounced with $d]_\sigma$
  - Intervocalically:
    - /t/ produced more accurately than /d/ for both age groups

	bi-morphemic		monomorphemic	
	/t/	/d/	/t/	/d/
Age 2;6	83%	5%	100%	28%
Age 3;6	97%	13%	99%	75%

Acquisition of Voicing Alternation

- ZKF's results:
  - Children never produced [d] syllable-finally
    - /bɛd/ and /pɛt/ never pronounced with $d]_\sigma$
  - Intervocalically:
    - /t/ produced more accurately than /d/ for both age groups
    - Monomorphemic words more accurate than bi-morphemic
      - Both for /t/s and /d/s

	bi-morphemic		monomorphemic	
	/t/	/d/	/t/	/d/
Age 2;6	83%	5%	100%	28%
Age 3;6	97%	13%	99%	75%

(lots of) Questions

- How can we explain * d$]_\sigma$?
  - Seems like phonotactics…
- How can we explain /VdV/ → [VtV] when the target language allows [VdV]?
  - Seems like the typical Markedness » Faithfulness effect
  - But this seems contrary to phonotactics…
    - How can the grammar capture what's permissible and at the same time perform unfaithful mappings?
- What explains the low accuracy on bi-morphemic /d/ relative to mono-morphemic /d/?
  - Bi-morphemic /d/ alternates with [t]…
- What explains the low accuracy of bi-morphemic /t/ relative to mono-morphemic /t/?
  - But neither bi-morphemic /t/ nor mono-morphemic /t/ alternates with [d]…
- What evidence is there for the role of phonotactics?

How can we make sense of all this?

- Computational Modeling
  - I will show that a computational model exhibits the same effects given data representative of the Dutch distribution
  - The model provides a possible explanation of the experimental findings
- Next
  - The learning theory
  - Dutch simulation and results
  - Implications and discussion

The Learning Theory: Overview

- Learning is formalized as optimization in MLG (Maximum Likelihood Learning of Lexicons and Grammars; Jarosz 2006).
  - Learning is gradual *likelihood maximization*
    - Finding the grammar and lexicon *combination* that best fits the data
- Optimization in Two Stages:
  - Phonotactic Learning
    - No morphological awareness
    - Learning of legal, illegal, and preferred phonotactics
  - Morphophonemic Learning
    - Words are analyzed into component morphemes
    - Learning of morpheme-specific underlying forms occurs
    - Further gradual learning of the grammar to account for alternations

The Learning Theory: Stages

- Phonotactic Learning
  - Find the grammar that maximizes likelihood given a *fixed rich base*
    - The rich base is simply the space of possible underlying forms
    - Held constant - no learning of URs is happening
  - This grammar encodes the biases/restrictions of the target language
- Morphophonemic Learning
  - Initialize with the phonotactic grammar…
  - Gradually converge on the grammar and lexicon combination that maximizes likelihood
    - Each morpheme begins with an unbiased (flat) distribution over all possible URs
      - Gradually, these distributions settle on the target URs
    - Further learning of the grammar to account for unfaithful mappings
  - Production hypothesized to begin at the onset of this stage
    - Production possible once morphemes get separate URs

The Learning Theory: Example

- Consider:
  - Data: 1 - [rat]$_{1}$ 1- [rade]$_{1, 3}$ 3 - [rat]$_{2}$ 3 - [rate]$_{2, 3}$
  - Constraints: Ident, NoSFV, NoVoi

The Learning Theory: Example

- Consider:
  - Data: 1 - [rat]$_1$ 1- [rade]$_{1, 3}$ 3 - [rat]$_2$ 3 - [rate]$_{2, 3}$
  - Constraints: Ident, NoSFV, NoVoi
- Phonotactic Learning:
  - 25% - /rad/ [rad] - 0 (0%)
  - 25% - /rat/ [rat] - 4 (50%)
  - 25% - /rate/ [rate] - 3 (37.5%)
  - 25% - /rade/ [rade] - 1 (12.5%)

The Learning Theory: Example

- Consider:
  - Data: 1 - [rat]$_{1}$ 1- [rade]$_{1, 3}$ 3 - [rat]$_{2}$ 3 - [rate]$_{2, 3}$
  - Constraints: Ident, NoSFV, NoVoi
- Phonotactic Learning:
  - 25% - /rad/ [rad] - 0 (0%)
  - 25% - /rat/ [rat] - 4 (50%)
  - 25% - /rate/ [rate] - 3 (37.5%)
  - 25% - /rade/ [rade] - 1 (12.5%)

The Learning Theory: Example

- Consider:
  - Data: 1 - [rat]$_1$ 1- [rade]$_{1,3}$ 3 - [rat]$_2$ 3 - [rate]$_{2,3}$
  - Constraints: Ident, NoSFV, NoVoi
- Phonotactic Learning:
  - 25% - /rad/ [rad] - 0 (0%)
  - 25% - /rat/ [rat] - 4 (50%)
  - 25% - /rate/ [rate] - 3 (37.5%)
  - 25% - /rade/ [rade] - 1 (12.5%)
- Morphophonemic Learning:
  - [ra**t**]$_1$ ~ [ra**d**e]$_{1,3}$ (/rate/ → [rate]) ⇒ */rat/$_1$
  - [ra**t**]$_2$ ~ [ra**t**e]$_{2,3}$ Pr(/ra**t**e/→[ra**t**e]) > Pr(/ra**d**e/→[ra**t**e]) ⇒ /rat/$_2$ > /rad/$_2$
  - Meanwhile… /rad/$_1$ ⇒ Ident » NoVoic

Implementation

- MLG formally defines learning at an abstract level:
  - Grammar is a generative probability model
  - Lexicon is a generative probability model
  - Learning is optimization
- In the simulations reported here, I make the following simplifying assumptions:
  - A grammar is a list of rankings with associated probabilities
  - A lexicon is a list of underlying forms with associated probabilities
  - Optimization occurs via the Expectation Maximization algorithm (Dempster et al 1977)
- In current work, I am exploring alternative implementations
- But the focus here is on what explanation this theory might offer:
  - EM is gradual and provably convergent
  - Sufficient to illustrate the predictions of the theory

Input: data distribution

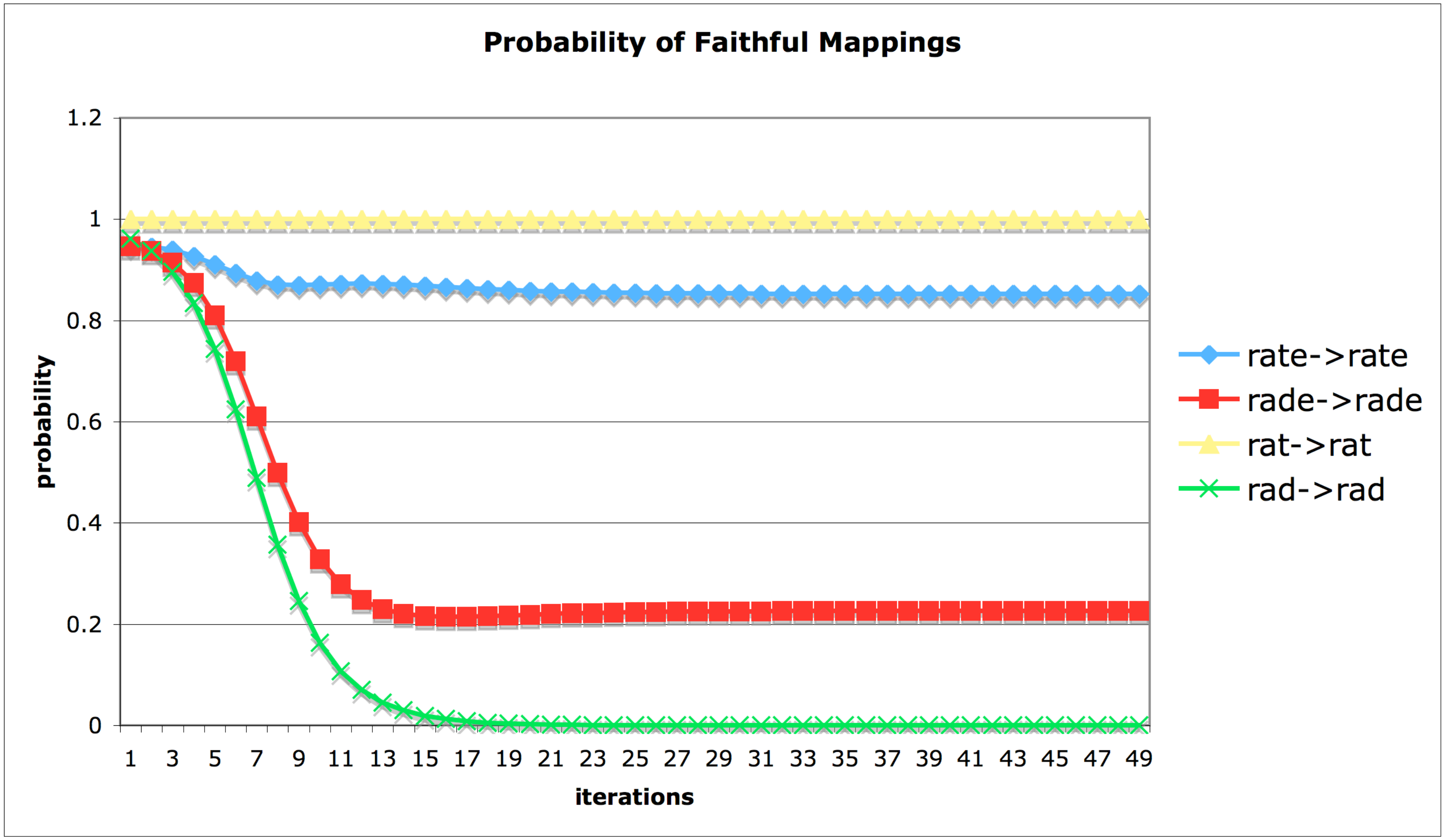
- Dutch data distribution provided to the model
- Frequencies below as reported by ZKF:
  - Overall: [t] - 95.1%, [d] - 4.9% overall
  - Intervocalically: [t] - 85%, [d] - 15%
  - Alternations: /d/ → [d] 0.5% overall

/t/→[t] *(sg.)*	*/t/→[t]* *(pl.)*	*/d/→[t]* *(sg.)*	*/d/→[d]* *(pl.)*	*monomorph. /t/*	*monomorph. /d/*
120 (11.8%)	**87 (8.6%)**	**564 (55.5%)**	**5 (0.5%)**	**196 (19.3%)**	**45 (4.4%)**
[rat_1]	**[$rat_1 + e_5$]**	**[rat_2]**	**[$rad_2 + e_5$]**	**[$rate_3$]**	**[$rade_4$]**

Input: constraints & base

- Simulation relies on the following constraints:
  - **NoVoi** - no voiced obstruents
  - **NoSFV** - no syllable-final voiced obstruents
  - ***VTV** - no voiceless obstruents intervocalically
  - **Ident[vc]** - no changes in voicing specification
  - **Max** - no deletion
- Note:
  - Target grammar: NoSFV, Max » Ident[vc] » NoVoi, *VTV
  - Given these constraints, /t/ or /d/ are possible URs in either context
- Space of URs provided to the model:
  - /rat/
  - /rad/
  - /rate/
  - /rade/

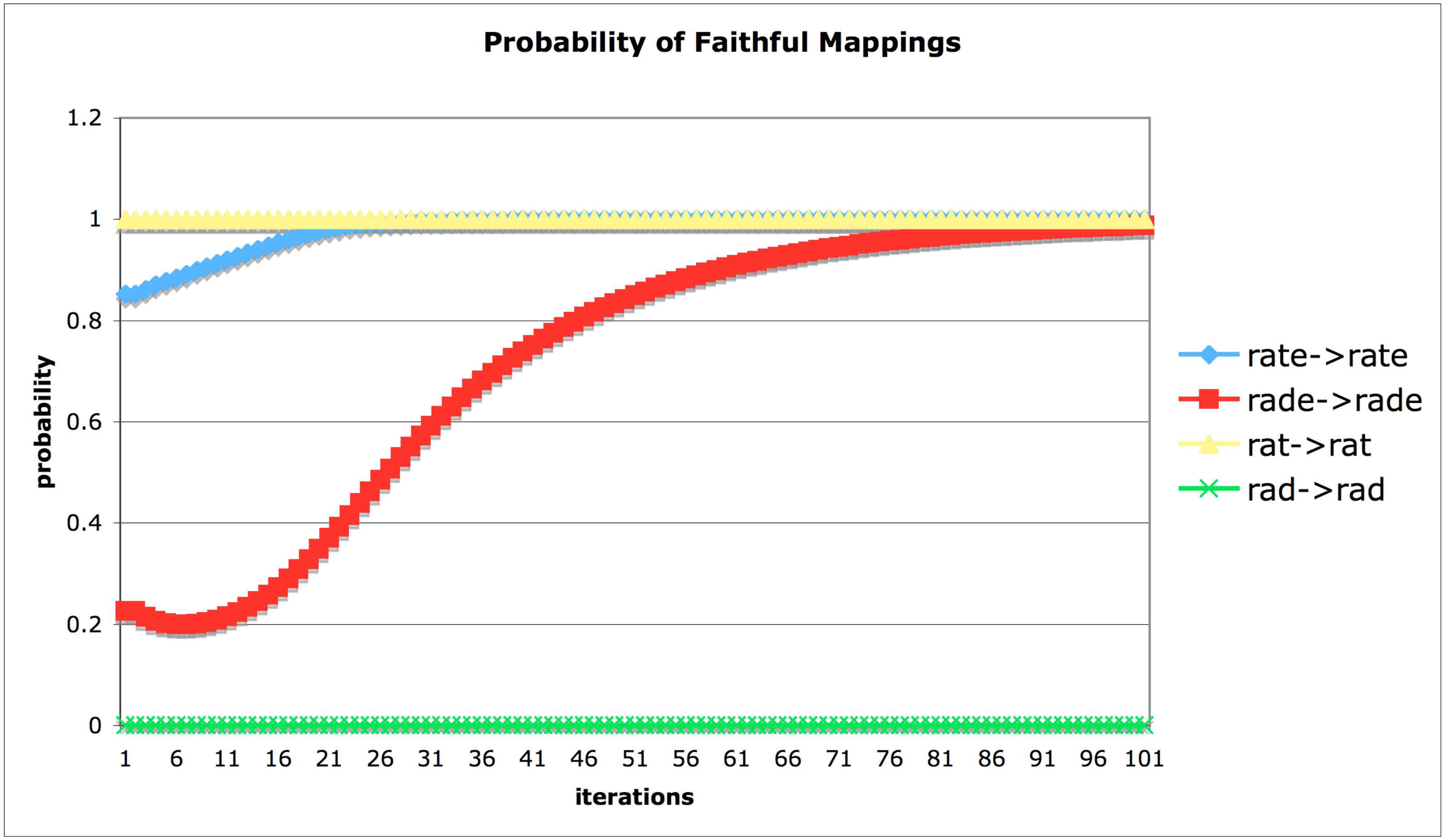
Simulation: Phonotactics



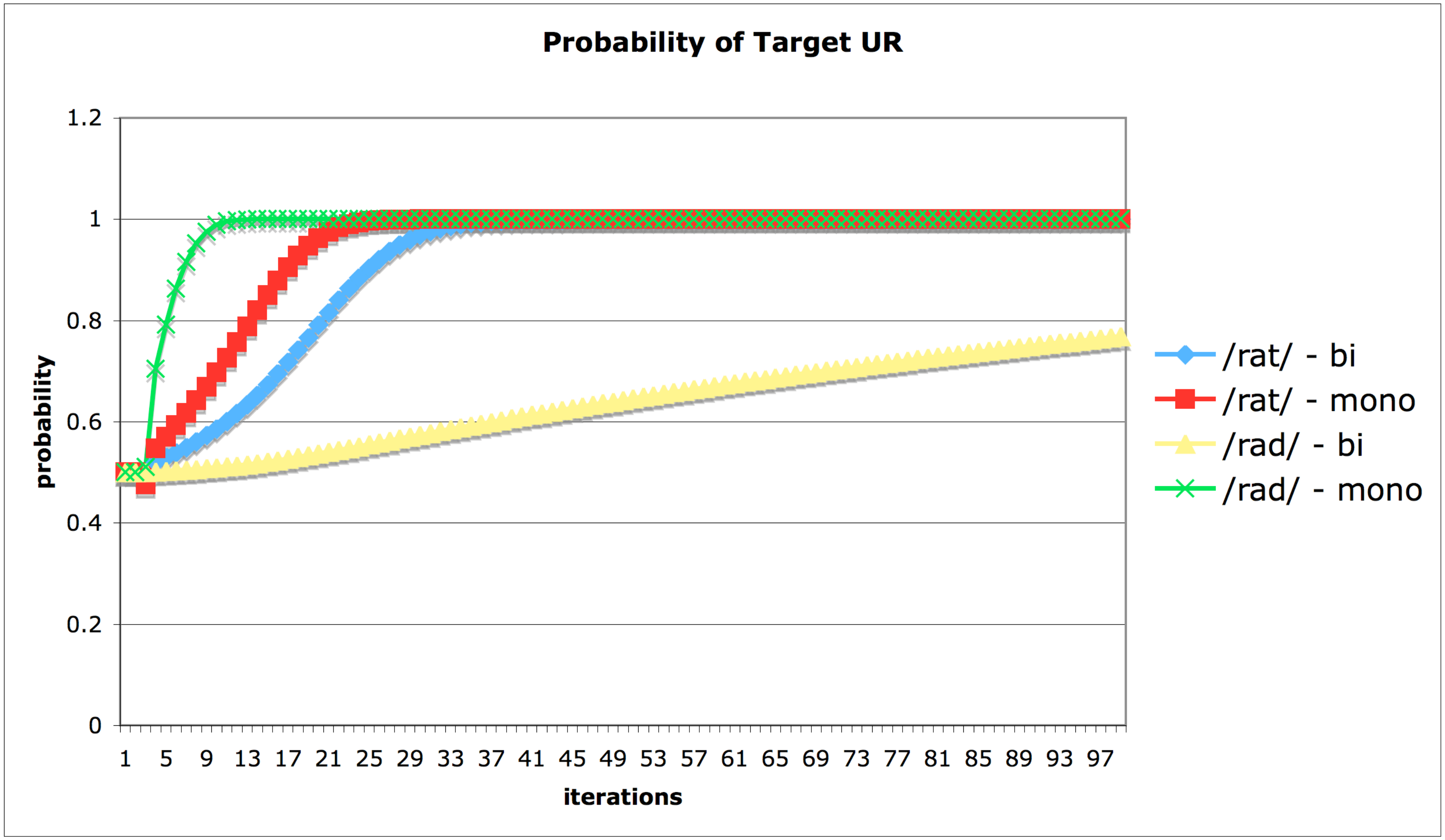
Simulation: Phonotactics

- The learned phonotactic grammar:
  - Voiced obstruents *not permitted* in coda:
    - Pr(/rad/ → [rad]) = 0
  - Both voiced and voiceless *permitted* intervocalically:
    - Pr(/ra**d**e/ → [ra**d**e]) > 0
    - Pr(/ra**t**e/ → [ra**t**e]) > 0
  - But *preference* for voiceless intervocalically:
    - Pr(/ra**d**e/ → [ra**t**e]) = 77.4%
    - ***Voi** probabilistically above **Ident**

Simulation: Morphophonemics



Simulation: Morphophonemics



Simulation: Production Accuracy

- The model:

	bi-morphemic		monomorphemic	
	/t/	/d/	/t/	/d/
Iter. 112	87%	15%	90%	23%
Iter. 126	96%	29%	99%	51%

- The experiment:

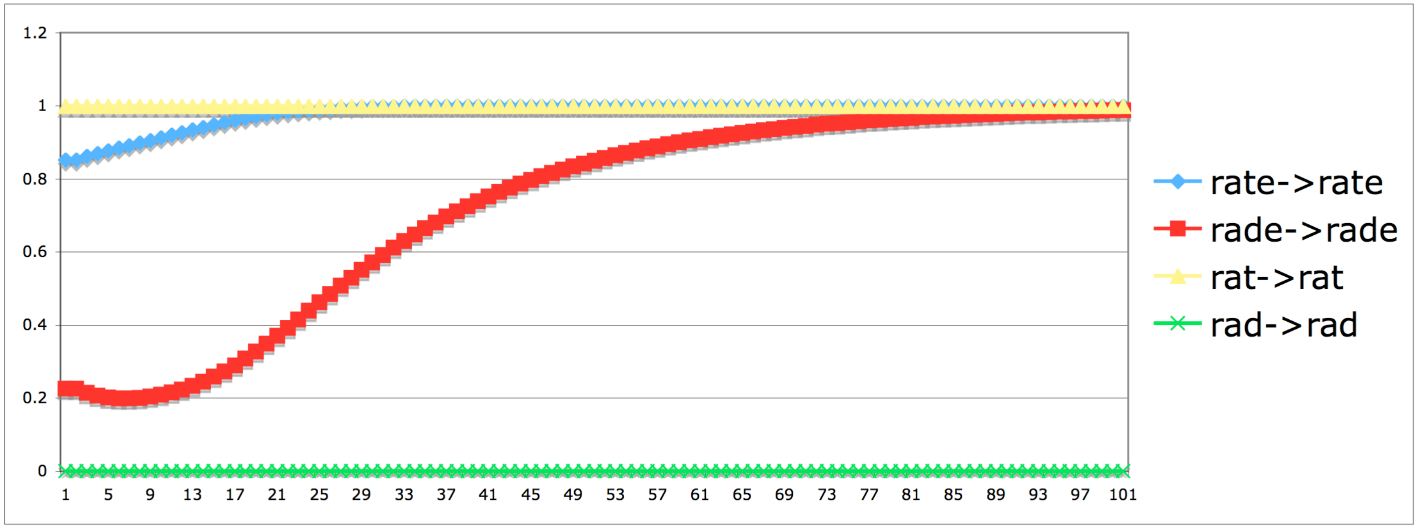
	bi-morphemic		monomorphemic	
	/t/	/d/	/t/	/d/
Age 2;6	83%	5%	100%	28%
Age 3;6	97%	13%	99%	75%

- Captured effects:
  - No voiced syllable-final obstruents (not shown)
  - [t] more accurate, especially early on
  - Mono-morphemic /d/ more accurate than bi-morphemic /d/
  - Mono-morphemic /t/ more accurate than bi-morphemic /t/

- Why do we see these effects in the model?

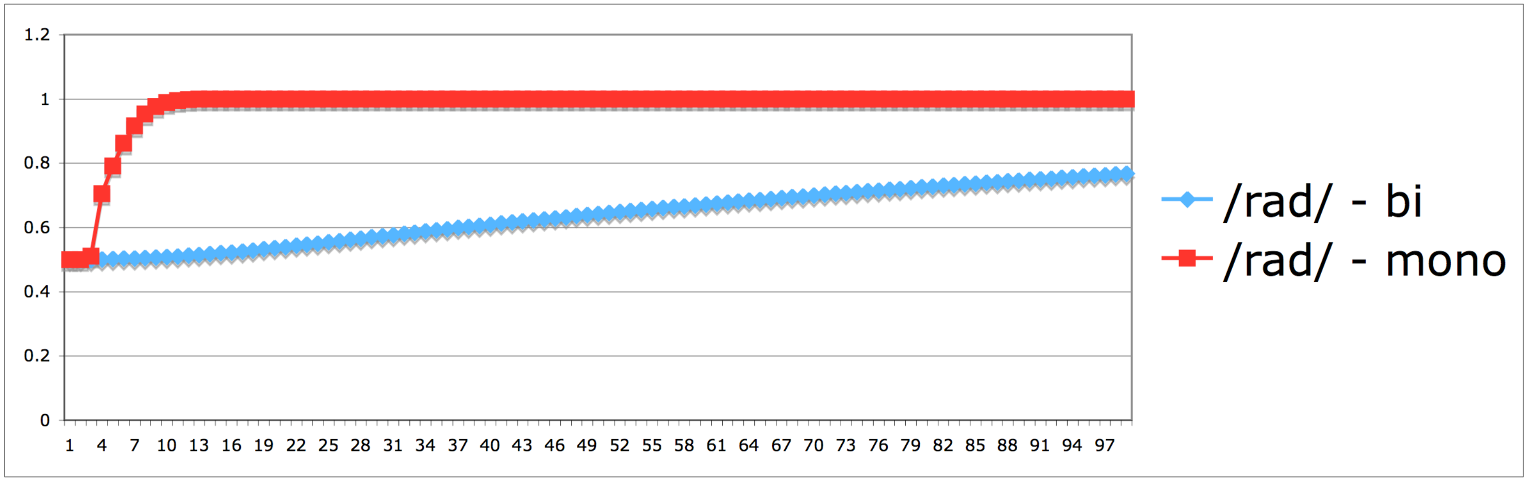
The Explanation: Phonotactics

- Phonotactics & frequency sensitivity explains:
  - No voiced syllable-final obstruents
    - Phonotactic learning ensures rankings permitting d]$_\sigma$ are eliminated
    - This ensures /d/ never surfaces surfaces faithfully
  - [t] more accurate, especially early on
    - phonotactic grammar captures the language-specific bias against [VdV]
- These are global, *grammatical* effects



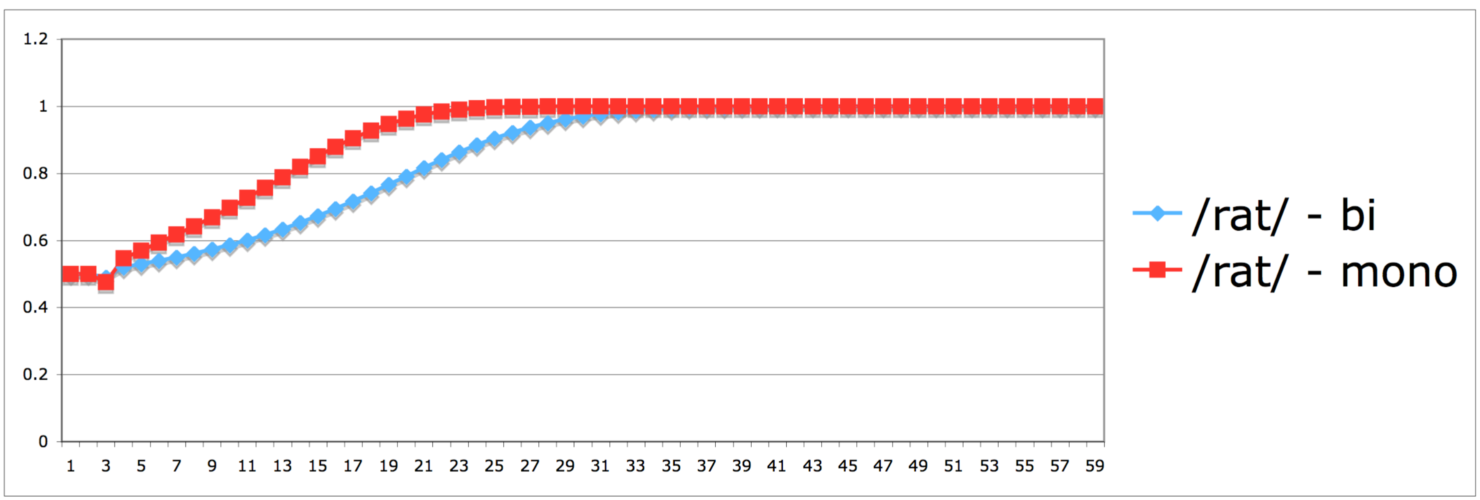
The Explanation: /d/

- Alternations, phonotactics & frequency explain:
  - Mono-morphemic /d/ more accurate than bi-morphemic /d/
    - Mono-morphemic /d/s occur intervocalically as [d] 100% of the time
      - Because of phonotactic knowledge intervocalic context provides evidence of /d/
    - Bi-morphemic /d/ occurs syllable-finally 99% of the time
      - Because of phonotactic knowledge, syllable-final context is equally consistent with /d/ and /t/
  - It is this *lexical delay* that results in lower accuracy for alternating /d/
    - Learner is unsure about the underlying voicing for alternating /d/ longer…
      - The difference in accuracy between the /d/s results from different URs
    - Relative frequency & informativeness of contexts in which a morpheme occurs



The Explanation: /t/

- Phonotactics and Frequency explain:
  - Mono-morphemic /t/ occurs in intervocalic context 100% of the time
  - Bi-morphemic /t/ occurs in syllable final position 58% of the time
    - Because of phonotactic knowledge, syllable final position is not informative
- Most direct evidence of phonotactic knowledge:
  - /t/ is always realized as [t]
  - Overall frequency of mono-morphemic /t/ is 19.3%
  - Overall frequency of bi-morphemic /t/ is 20.4%



Summary: General Predictions

- Effects of Phonotactics
  - Learning a setting for a feature in a contrastive context should be faster than in a neutralized context
  - Precisely because knowledge of phonotactics underdetermines the UR in neutralizing contexts
- Relative frequency & alternation effects
  - The rate of UR learning *for a given morpheme* depends on the proportion of surface realizations that are informative about URs
    - For Dutch: URs for stems with a higher relative proportion of plurals should be acquired more quickly
- Probabilistic phonotactics effects
  - To the extent allowable by the constraint set, the phonotactic grammar will reflect statistical biases in the language
    - These biases will be apparent in initial productions

Discussion

- Understanding the properties of the model provides a possible explanation of these experimental findings
- Additionally, manipulating the input to the model makes testable predictions about learners' behavior in other languages/domains
  - Acquisition data can guide the refinement of the model
  - This could guide the development of future experiments

Necessary Properties

- Crucial aspects of the data distribution:
  - Low frequency of [VdV]
  - The fact that bi-morphemic /t/s and /d/s occur in syllable final context
- Sensitivity to frequency
  - Relative frequency of alternants & their contexts affects rate of learning
  - Also, phonotactic biases are sensitive to statistical biases
- Lexical representations that are ‘rich’ even for non-alternating morphemes
  - If both /t/s started as /t/ no difference in accuracy would be expected
- Morphophonemic learning starts from a phonotactic grammar that:
  - Encodes language-specific phonotactic restrictions,
    - M >> F predicts initial production is sensitive only to *universal* markedness
  - Penalizes licit, marked forms
    - If we just had Ident >> *Voi, *VTV, initial productions would be perfect in contrastive contexts

Outstanding Questions

- Can we *derive* (i.e. explain) phonotactics $\preccurlyeq$ alternations from some independently motivated criteria?
  - And of course this division is unlikely to be quite so perfect
- What is the role of learning the morphological associations?
  - In a sense, the model predicts phonotactics *hurts* learning URs
  - But phonotactics should be very helpful for learning this aspect of alternations
- What is the development of voicing alternations in languages with:
  - a higher proportion of intervocalic contexts
  - no/less preference for voicelessness overall

The End

- Thank You!

Selected References

- Jarosz, Gaja. 2006. Rich Lexicons and Restrictive Grammars - Maximum Likelihood Learning in Optimality Theory. PhD dissertation, Johns Hopkins University.
- Zamuner, Tania S., Annemarie Kerkhoff, & Paula Fikkert. In preparation. 'Children's knowledge of how phonotactics and morphology interact'.