EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Philosophische Fakultät
Seminar für Sprachwissenschaft

NDRa: a single route model of reading aloud based on discriminative learning

Peter Hendrix, Michael Ramscar & Harald Baayen

Outline

- Introduction
- NDRa model
- Simulations
    - Overall model fit
    - Predictor simulations
    - Comparison to dual-route architecture
    - Pronunciation performance
- Conclusions

Outline

- Introduction
- NDRa model
- Simulations
    - Overall model fit
    - Predictor simulations
    - Comparison to dual-route architecture
    - Pronunciation performance
- Conclusions

Introduction

- Existing models of reading aloud are dual-route models
- Lexical route
    - Orthography to phonology mapping is mediated by lexical representations
    - Responsible for reading known words (e.g.; *food*)
- Sub-lexical route
    - Direct orthography to phonology mapping
    - Responsible for reading unknown words (e.g.; *snood*)

Introduction

snood

Introduction

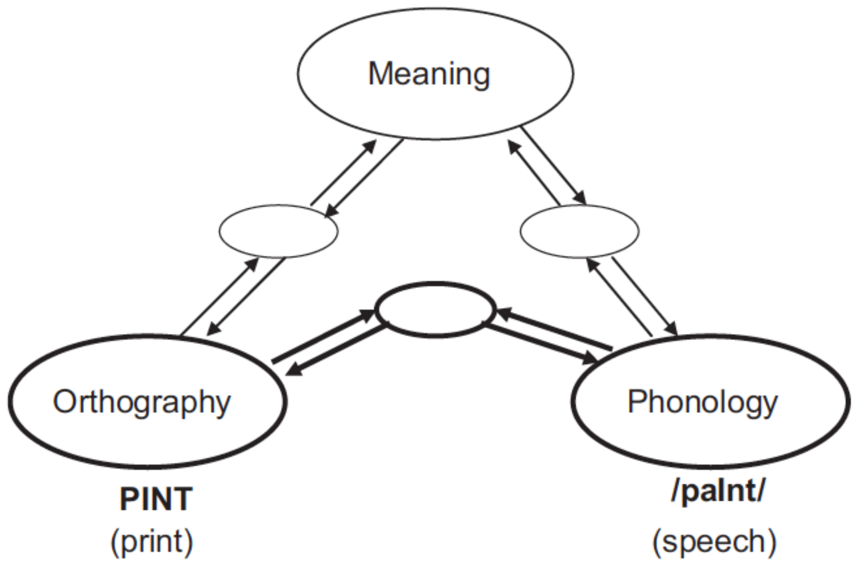
Introduction

- Examples of dual-route models:
    - Triangle model (Harm & Seidenberg, 2004)
    - DRC model (Coltheart et al., 2001)
    - CDP+ model (Perry et al., 2007)

Triangle model

- Connectionist model
- Three levels of description:
  - Orthography
  - Phonology
  - Semantics

Triangle model



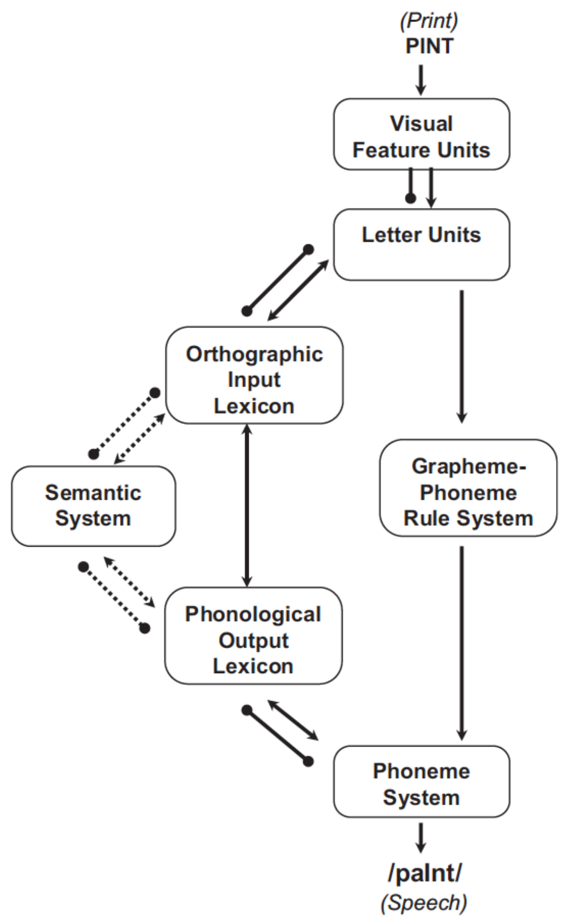
Triangle model

- What is represented by hidden layer units?
- Less explanatory power than CDP+ model

DRC model

- Lexical route: interactive activation model (McClelland & Rumelhart, 1981)
- Sub-lexical route: grapheme-to-phoneme conversion rules

DRC model


(Print)
PINT
Visual
Feature Units
Letter Units
Orthographic
Input
Lexicon
Semantic
System
Grapheme-
Phoneme
Rule System
Phonological
Output
Lexicon
Phoneme
System
/paInt/
(Speech)

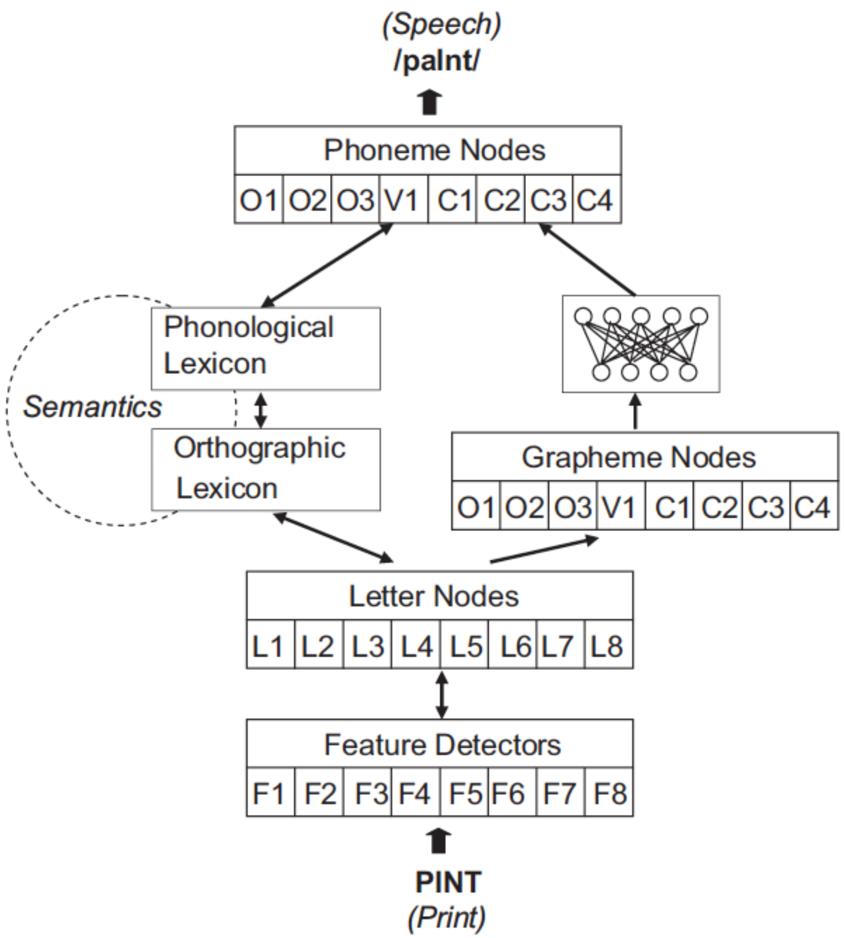
DRC model

- Ignores the problem of learning in both routes
- Poor performance compared to newer models of reading aloud

CDP+ model

- Successor of DRC model
- Hybrid model:
  - Lexical route: interactive activation model
  - Sub-lexical route: discriminative learning network (Zorzi et al., 1998)

CPD+ model


(Speech)
/paInt/
Phoneme Nodes
O1 O2 O3 V1 C1 C2 C3 C4
Phonological Lexicon
Semantics
Orthographic Lexicon
Grapheme Nodes
O1 O2 O3 V1 C1 C2 C3 C4
Letter Nodes
L1 L2 L3 L4 L5 L6 L7 L8
Feature Detectors
F1 F2 F3 F4 F5 F6 F7 F8
PINT
(Print)

CDP+ model

- Performs an order of magnitude better than other existing models of reading aloud
- Ignores the problem of learning in the lexical route

Introduction

- Can a learning network improve the performance of the lexical route?
- Is a sub-lexical route really necessary?

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

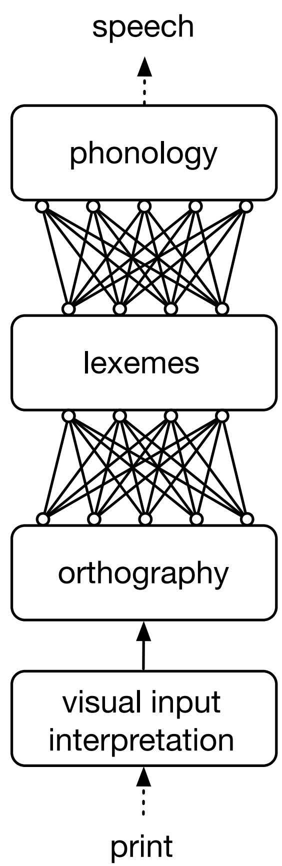
Outline

- Introduction
- NDRa model
- Simulations
    - Overall model fit
    - Predictor simulations
    - Comparison to dual-route architecture
    - Pronunciation performance
- Conclusions

NDRa

- Extension of the NDR model to reading aloud
- Single-route model: no sub-lexical route

NDRa model



speech
phonology
lexemes
orthography
visual input interpretation
print

NDRa

- Visual input interpretation based on Manhattan city-block measure (Han & Kamber, 2000)
- More complex visual patterns should take longer to decode
- Complexity of a letter is inversely proportional to the similarity of that letter to all other letters
- Complexity of a word is the summed complexity of all component letters

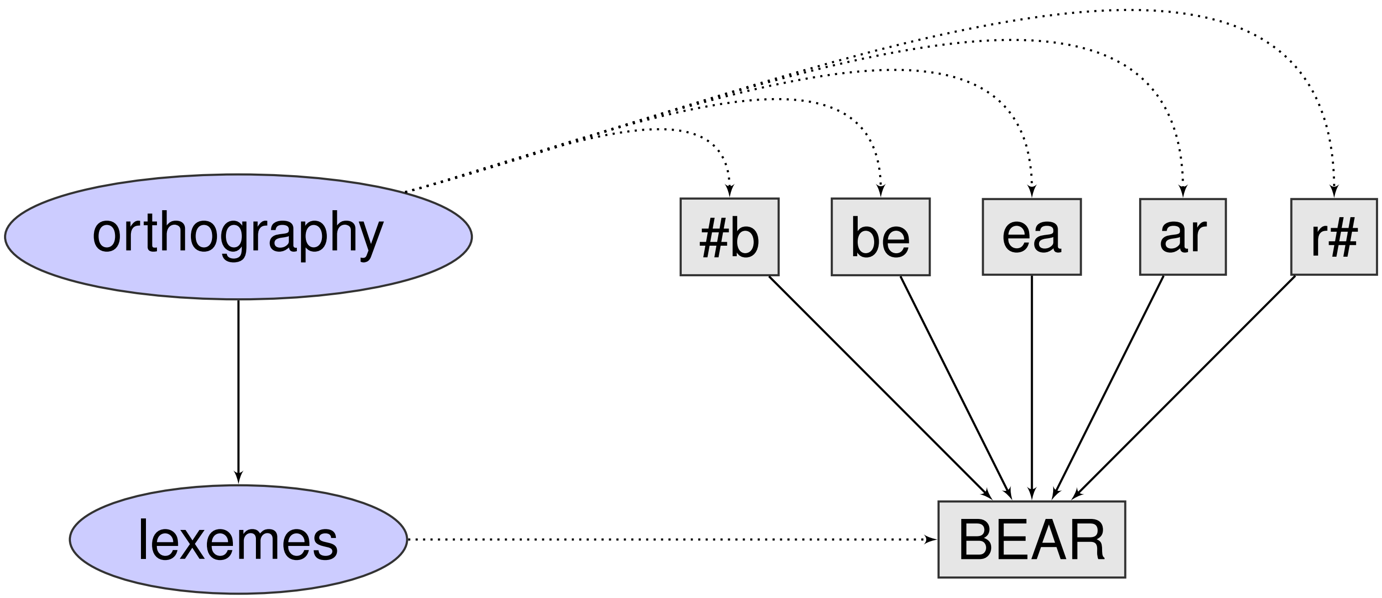
NDRa

- Two discriminative learning networks:
    - Orthography to lexemes
    - Lexemes to phonology

NDRa

- Orthography to lexeme network:
  - Original NDR model (Baayen et al., 2011)
  - Input units: letters and letter bigrams (e.g.; *#b, be, ea, ar, r#*)
  - Outcomes: lexemes (e.g.; *BEAR*)

NDRa

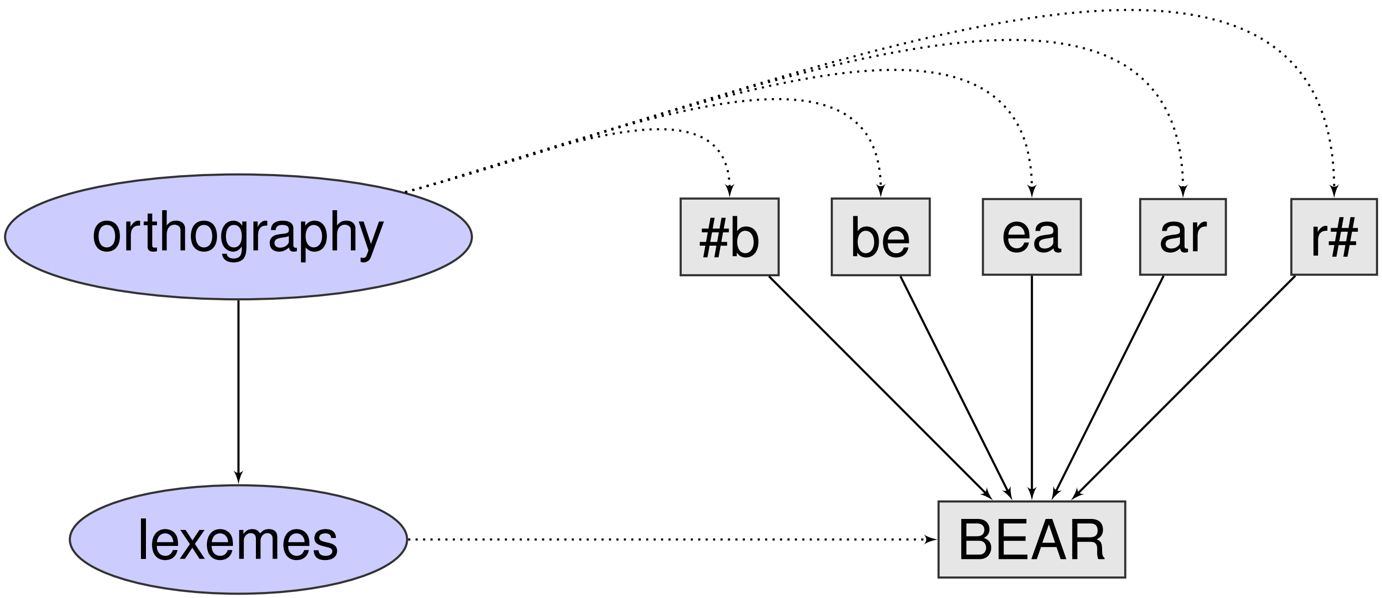


EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

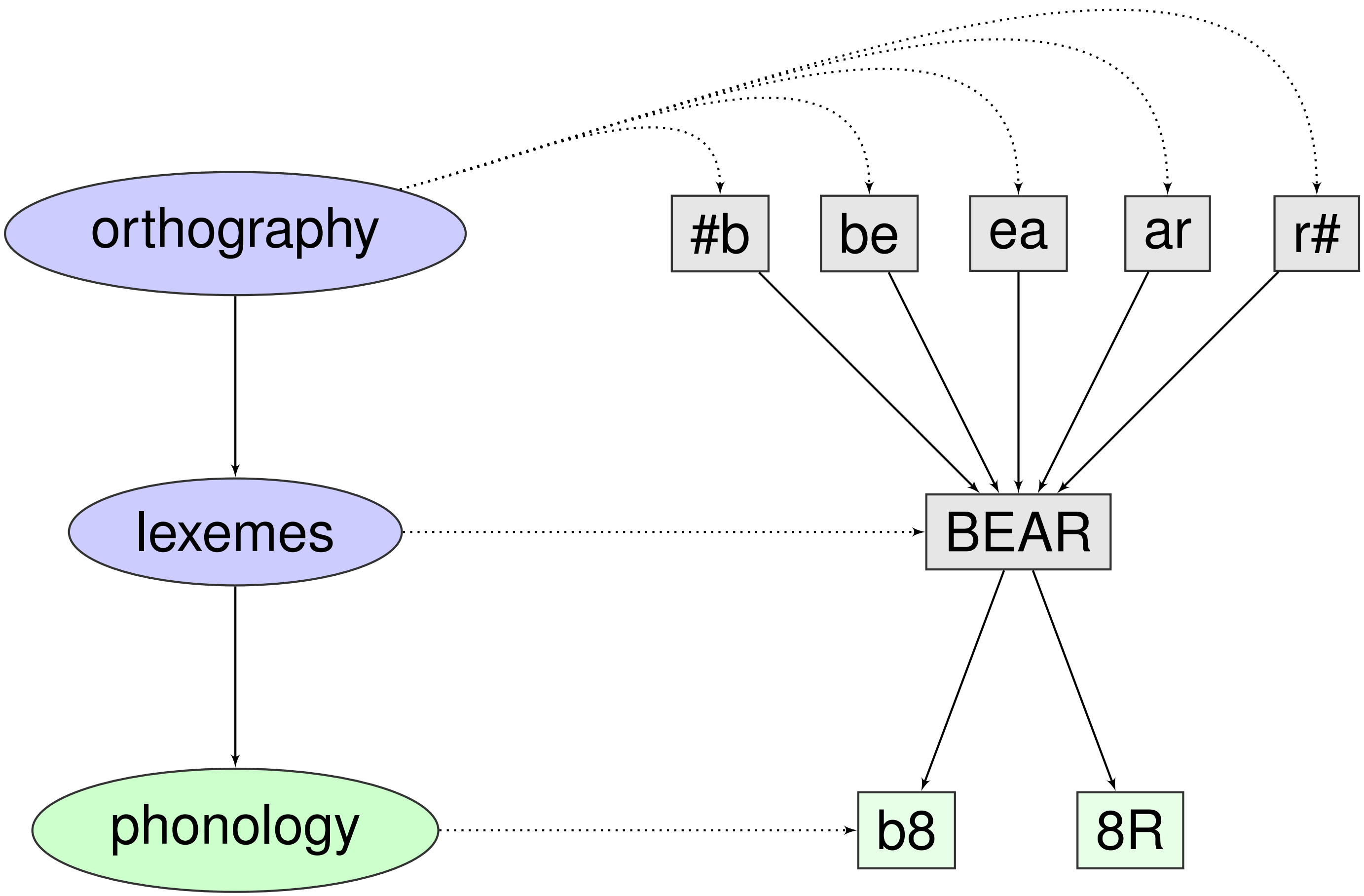
NDRa

- Lexeme to phonology network:
  - New in the NDRa model
  - Input units: lexemes (e.g.; *BEAR*)
  - Outcomes: demi-syllables (e.g.; *b8, 8r*)

NDRa



NDRa

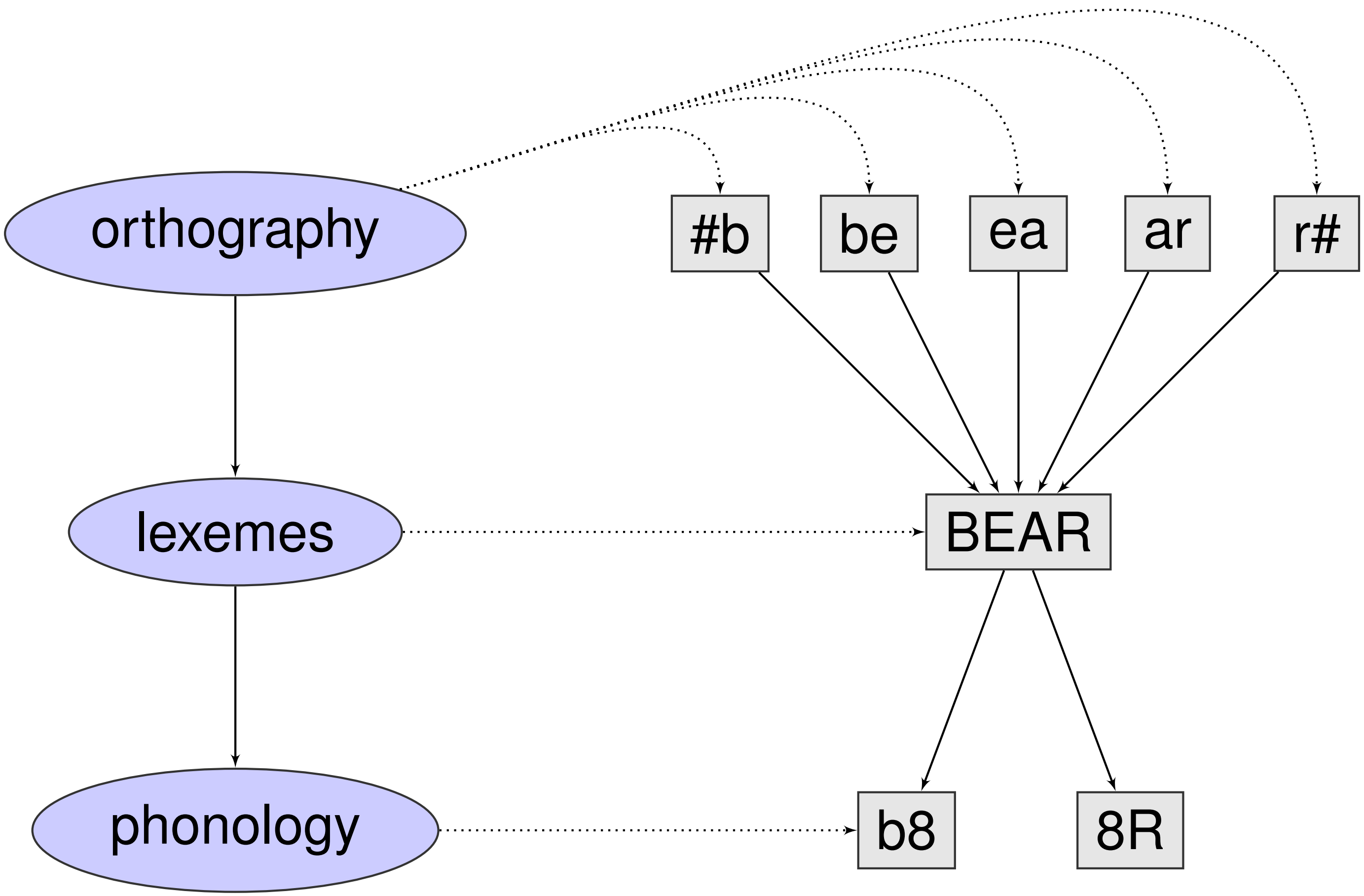


EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

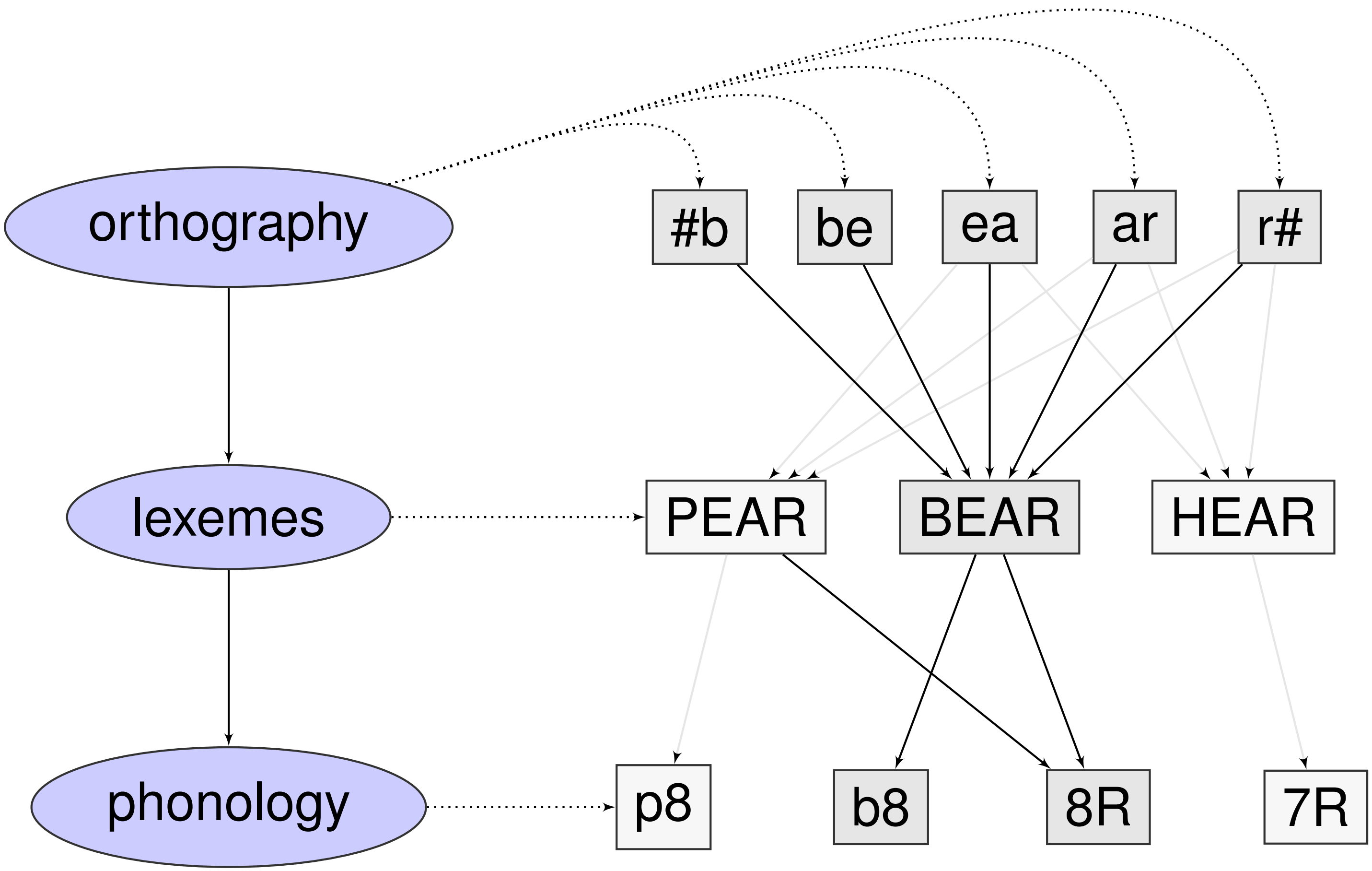
NDRa

- Orthographic units activate not only the lexeme of the target word, but also the lexemes of orthographic neighbors of the target word
- Phonological units are activated by the lexeme of the target word as well as by the lexemes of the activated orthographic neighbors

NDRa



NDRa


orthography
lexemes
phonology
#b
be
ea
ar
r#
PEAR
BEAR
HEAR
p8
b8
8R
7R

NDRa

- Given the activation a_t from the target lexemes and the activations $a_{1,\ldots,n}$ from the lexemes of co-activated orthographic neighbors, the total activation of a demi-syllable k is defined as:

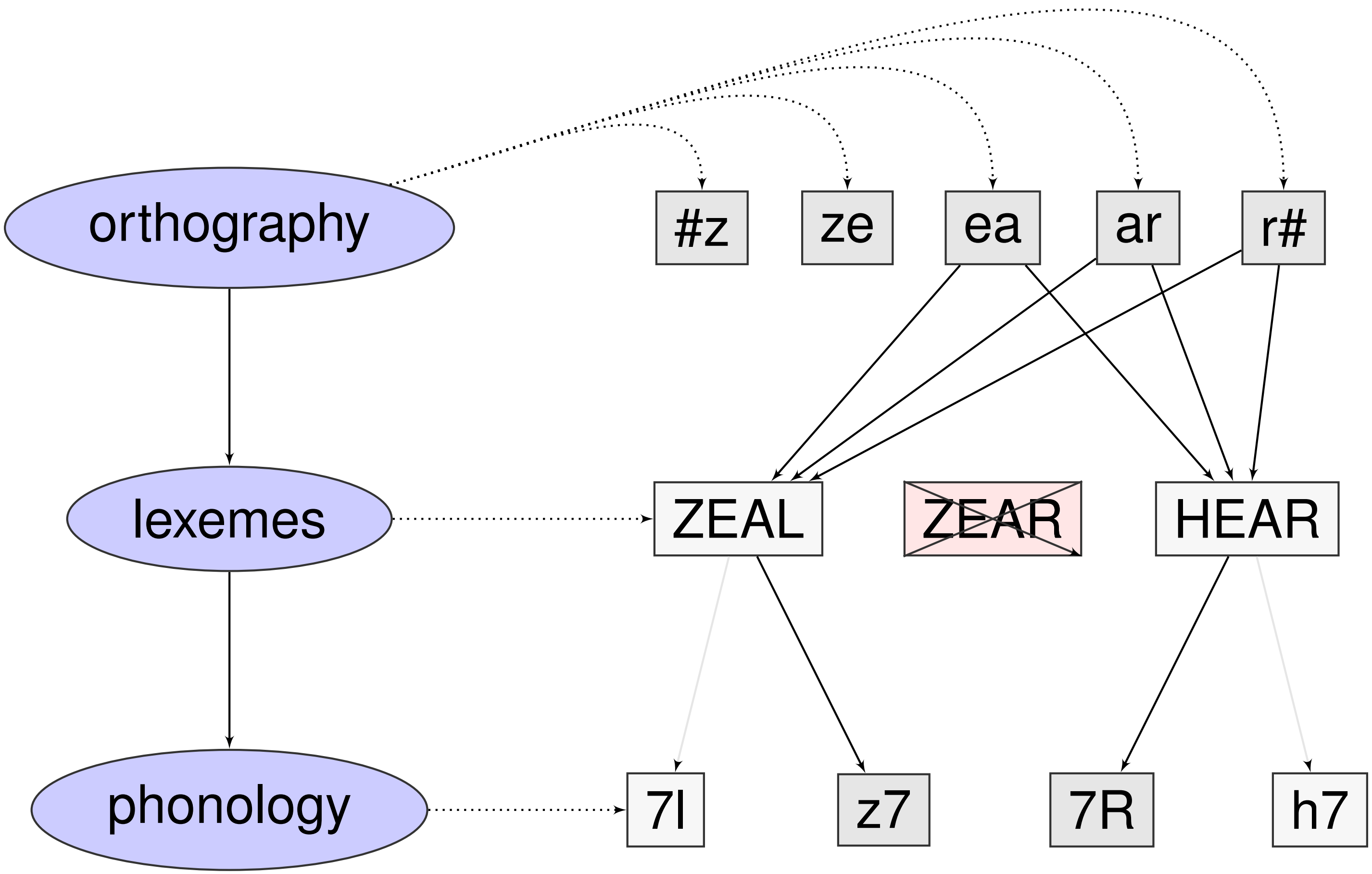
  $$ActPhon_k = a_t + \sum_{i=1}^{n} w_i * a_i$$

  where w_i is the amount of activation that the meaning of lexical neighbor i received from the orthography of the target word

NDRa

- Unknown words and non-words are processed by the same architecture
- No lexico-semantic representations exist for non-words
- Pronunciation is therefore mediated only by the activations of orthographic neighbors

NDRa


orthography
#z
ze
ea
ar
r#
lexemes
ZEAL
ZEAR
HEAR
phonology
7l
z7
7R
h7

NDRa

- NDRa models reading aloud of monosyllabic words
- Choice problem: which demi-syllable should be pronounced first?
- Modeled through the entropy over the activations of the first and second demi-syllable

NDRa

- Simulated reaction times are proportional to a weighted multiplicative integration of:
  - the complexity of the visual input
  - the activation of the target word lexeme
  - the activation of the demi-syllables of the target word
  - the entropy over the demi-syllable activations
- $RT \propto \frac{\text{Complexity}^{w1}}{\text{ActLexeme}^{w_2} * \text{ActPhon}_1^{w_3} * \text{ActPhon}_2^{w_4} * \text{H}^{w_5}}$

  where $w_{1,\ldots,5}$ are weight parameters that determine the relative contribution of each source of information

Outline

- Introduction
- NDRa model
- Simulations
    - Overall model fit
    - Predictor simulations
    - Comparison to dual-route architecture
    - Pronunciation performance
- Conclusions

Simulations

- 2416 mono-syllabic words
- 1697 non-words
    - 854 regular non-words
    - 843 pseudohomophones (e.g.; "bloo")
- 16 linguistic predictors

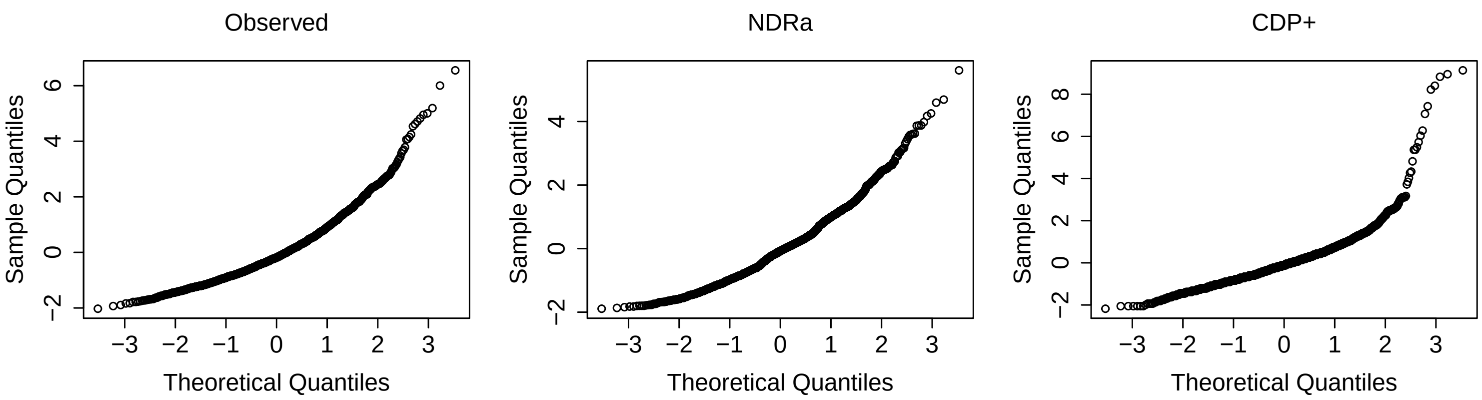
Outline

- Introduction
- NDRa model
- Simulations
  - Overall model fit
  - Predictor simulations
  - Comparison to dual-route architecture
  - Pronunciation performance
- Conclusions

Overall model fit

- Comparison of simulated latencies and observed ELP naming latencies:
  - $r = 0.48$ for both NDRa and CDP+
  - AIC much better for NDRa
  - Latency distribution much better for NDRa

Overall model fit



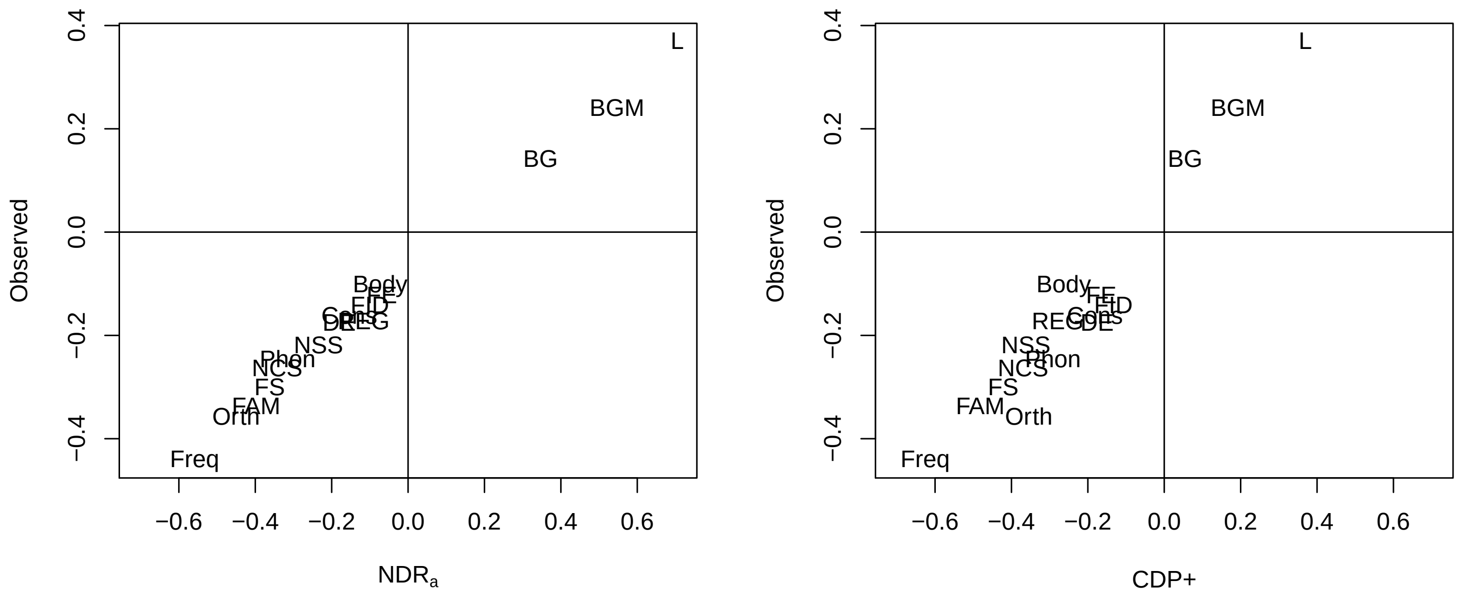
Outline

- Introduction
- NDRa model
- Simulations
    - Overall model fit
    - Predictor simulations
    - Comparison to dual-route architecture
    - Pronunciation performance
- Conclusions

Predictor simulations

- How well do both models capture the effects of the 16 linguistic predictors?
- Fit a separate linear model for each predictor
- Compare β coefficients between models for normalized observed and simulated latencies

Predictor simulations



L	=	length	FE	=	friends-enemies ratio	DE	=	derivational entropy	FS	=	family size
BGM	=	mean bigram freq.	FID	=	freq. initial diphone	NSS	=	number of simplex synsets	FAM	=	familiarity
BG	=	summed bigram freq.	Cons	=	consistency	Phon	=	phon. neighb. size	Orth	=	orth. neighb. size
Body	=	body neighb. size	REG	=	regularity	NCS	=	number of complex synsets	Freq	=	frequency

Predictor simulations

- Excellent performance for both models
- Nearly perfect correlation with observed β coefficients for the NDRa model ($r = 0.997$)
- CDP+ model seems to have problems with the relative contribution of neighborhood measures

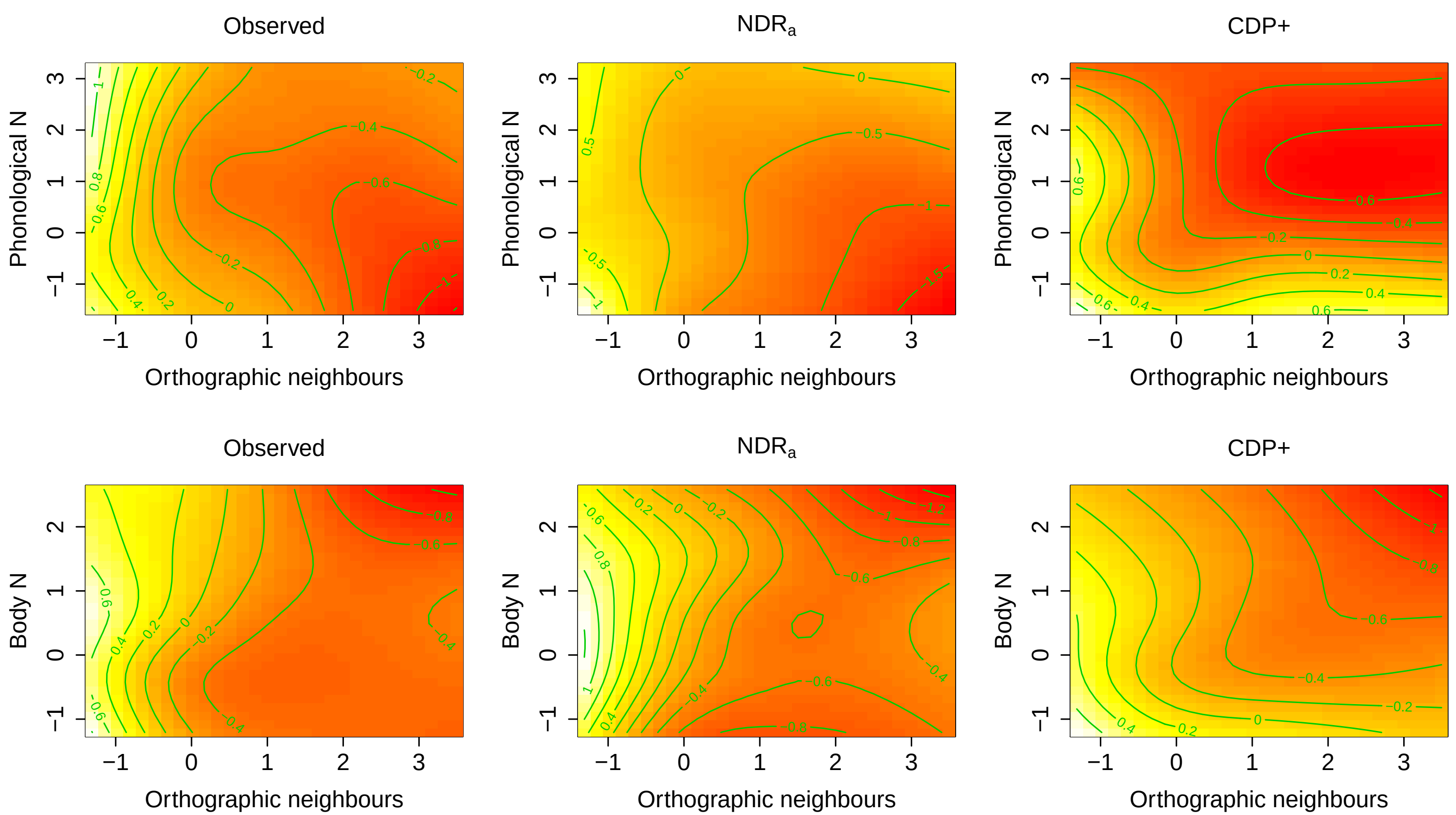
Predictor simulations: neighborhood measures

- Effects of three neighborhood measures have been documented:
  - Orthographic neighborhood (e.g.; *bear* - *pear*, *bear* - *hear*, *bear* - *bead*)
  - Phonological neighborhood (e.g.; *bear* - *pear*, *bear* - *hair*, *bear* - *bail*)
  - Body neighborhood (e.g.; *bear* - *pear*, *bear* - *wear*)

Predictor simulations: neighborhood measures

- Both models successfully capture the non-linear effects of all three predictors in isolation
- What about the non-linear interplay of the neighborhood measures?
- Find out using tensor product GAMs on simulated and observed latencies

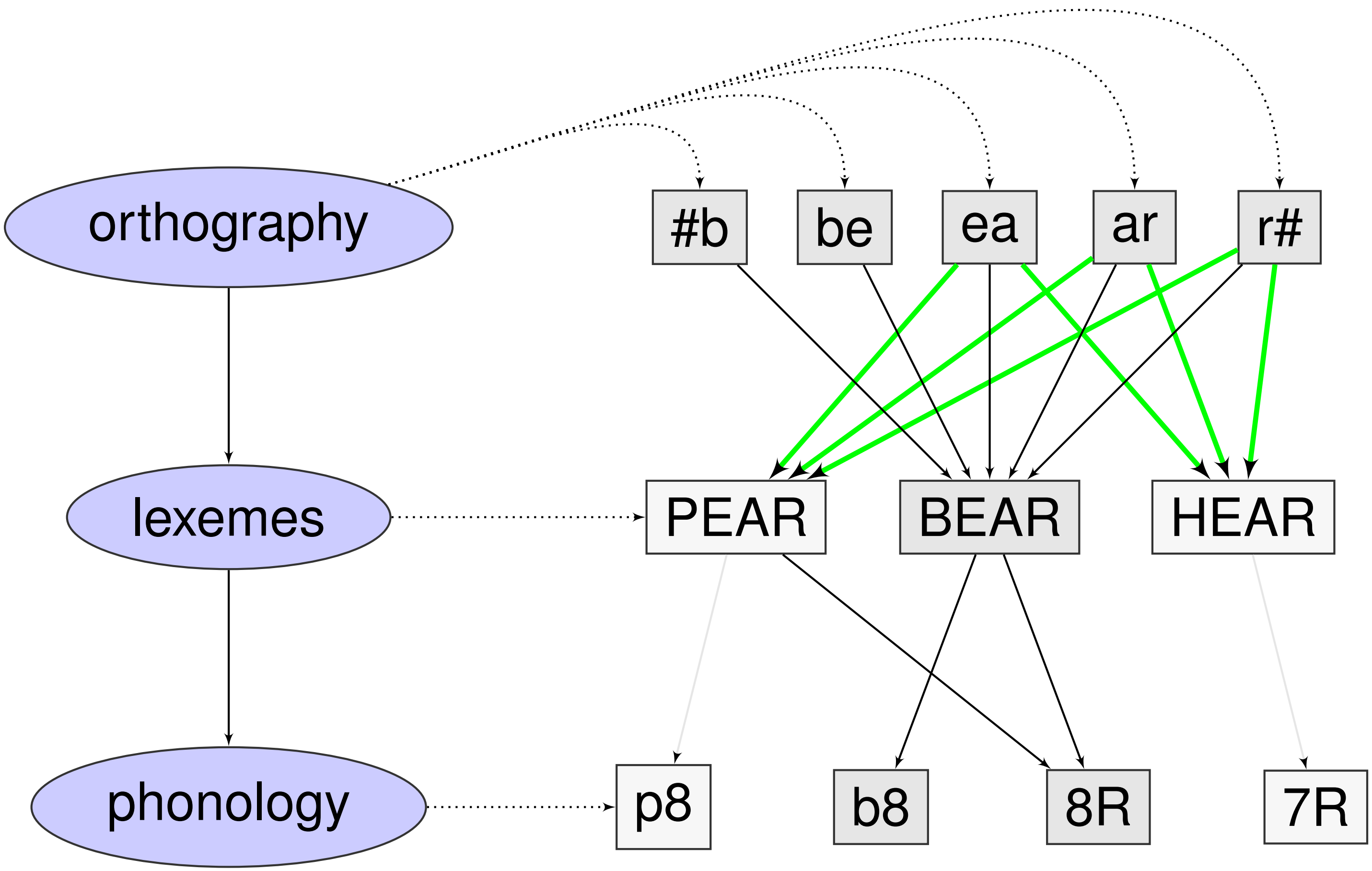
Predictor simulations



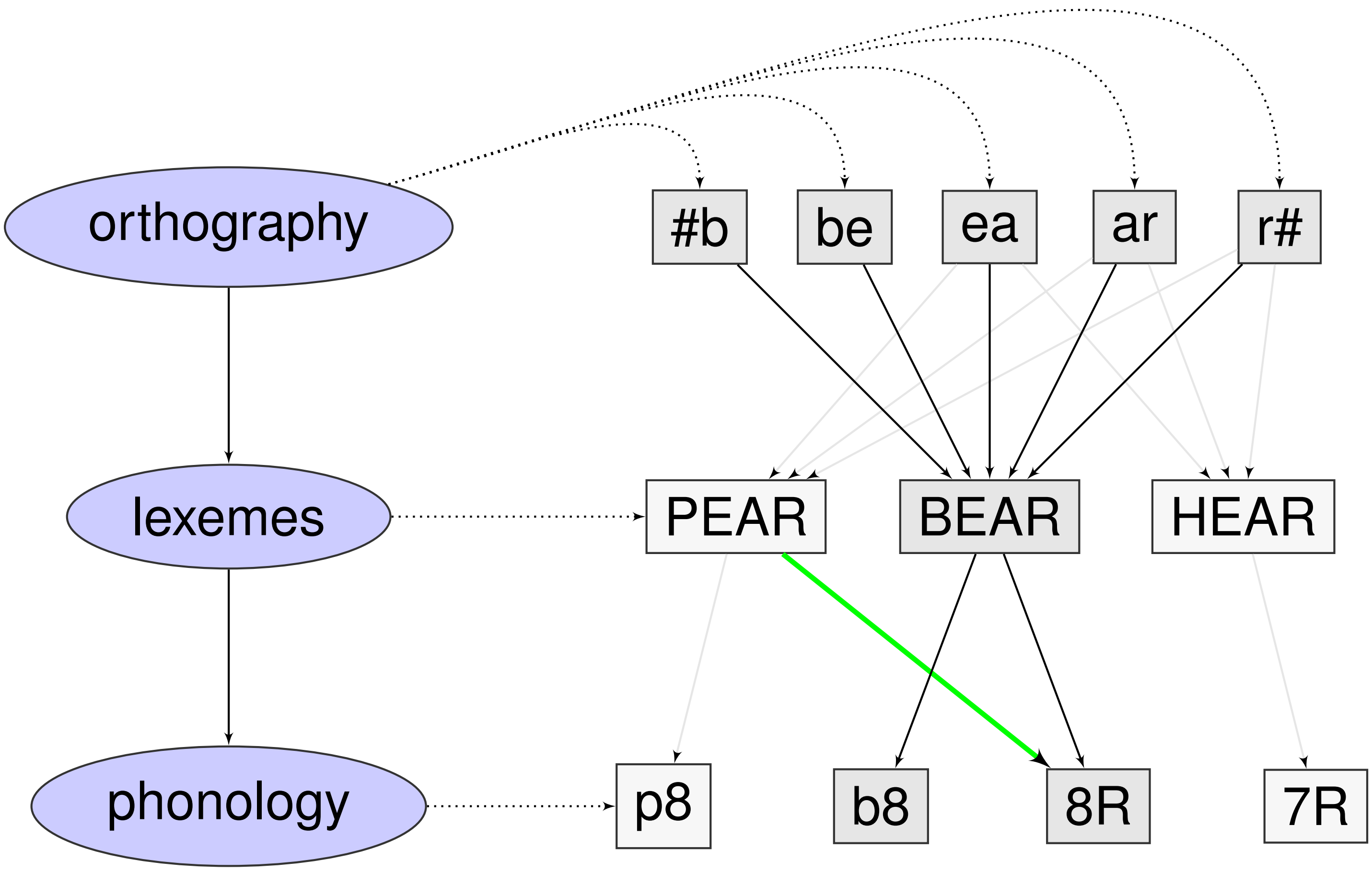
Predictor simulations: neighborhood measures

- Neighborhood effects are primarily orthographic neighborhood effects
- The NDR model correctly predicts the non-linear interplay of the neighborhood measures

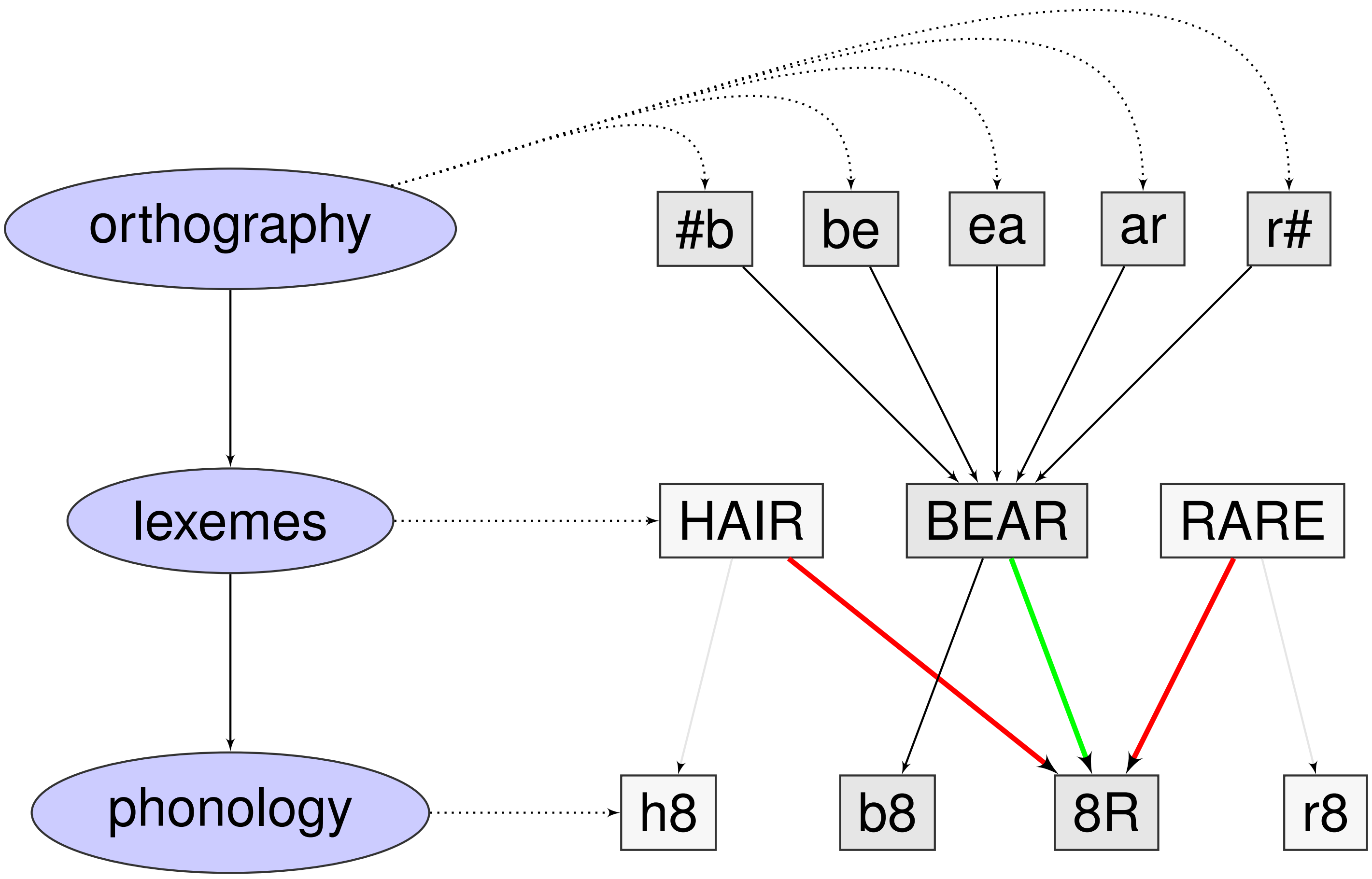
Predictor simulations: neighborhood measures



Predictor simulations: neighborhood measures



Predictor simulations: neighborhood measures



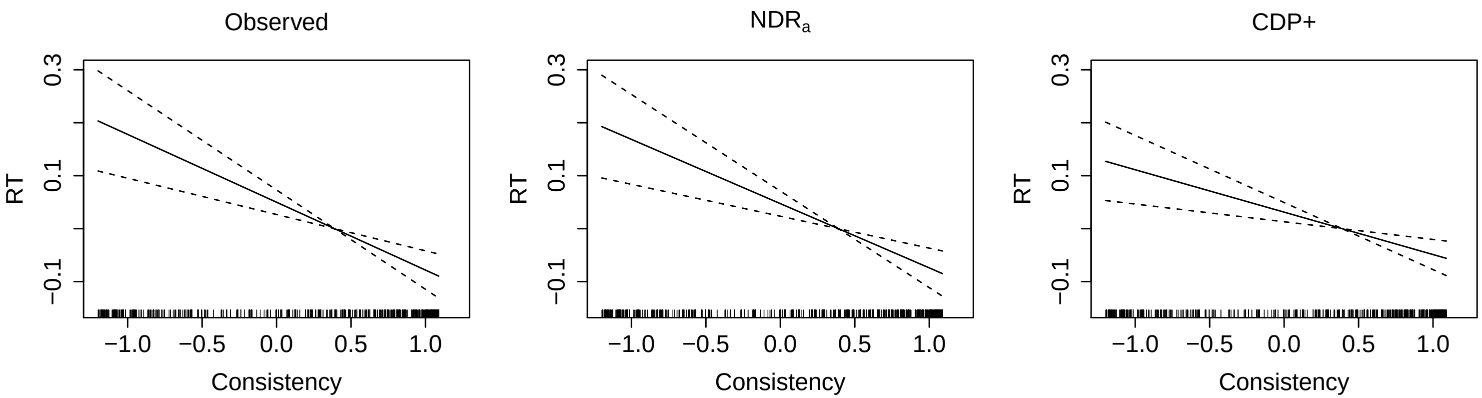
Predictor simulations: consistency

- The orthography to phonology mapping can be consistent or inconsistent
- Consistent with *pear*: *bear*, *wear*
- Inconsistent with *pear*: *dear*, *fear*, *gear*, *hear*, *lear*, *near*, *rear*, *year*, ...
- Higher proportions of consistent word tokens correspond to shorter naming latencies

Predictor simulations: consistency

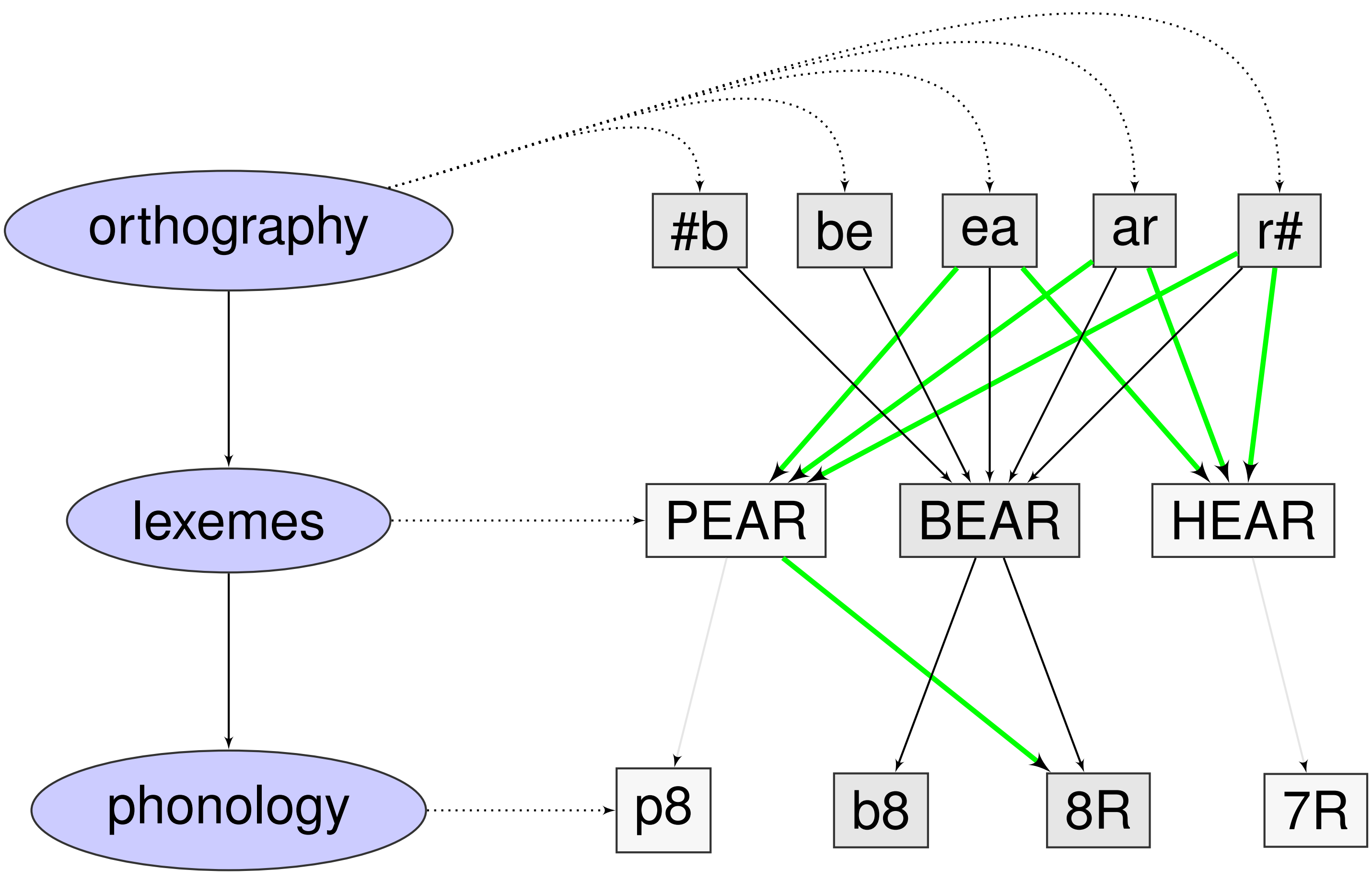
- Capturing consistency effects was a major advancement of the CPD+ model over the original DRC model
- In the CDP+ model consistency effects arise in the learning network in the sub-lexical route
- Can the single-route NDRa capture the effect of consistency?

Predictor simulations: consistency



EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Predictor simulations: consistency

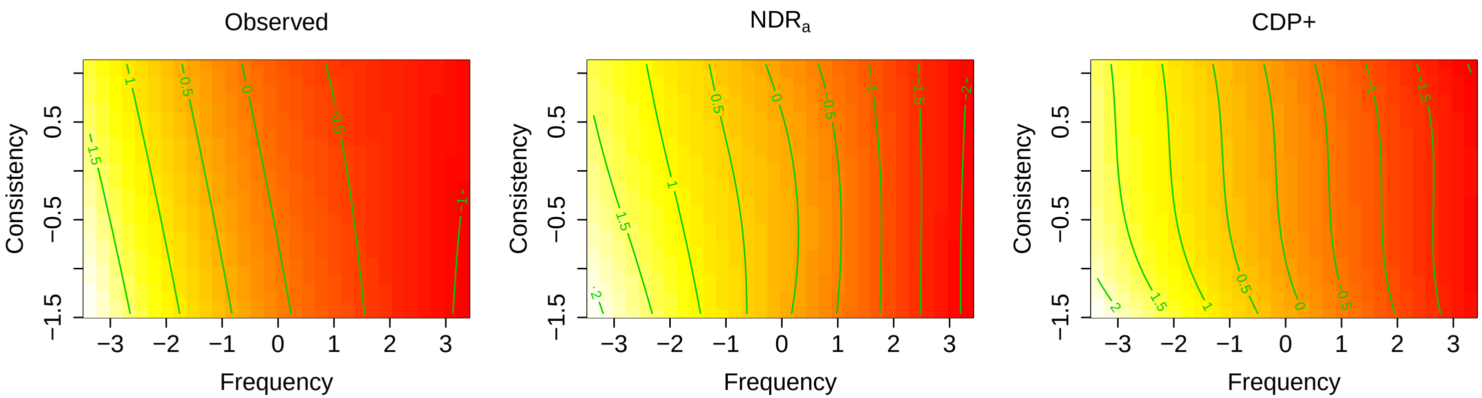


Predictor simulations: consistency

- Experimental data show a frequency by consistency interaction
- A consistency effect is observed for low frequency words only
- Does the NDRa capture this interaction?

Predictor simulations: consistency

Predictor simulations: consistency



Predictor simulations: non-words

- The NDRa successfully replicates a large number of effects in word naming
- How about non-word naming?

Predictor simulations: non-words

- Non-words naming effects captured by the NDRa include:
  - Non-words are read slower than real words
  - Non-word naming latencies increase linearly with length
  - Non-words with more orthographic neighbors are read faster
  - A higher proportion of consistent word tokens leads to shorter non-word naming latencies
  - Pseudohomophones are read faster than regular nonwords

Predictor simulations: non-words

- [Ramscar]Does the frequency of non-words help predict naming latencies?[/Ramscar]
- Re-analysis of naming latencies for non-words in McCann & Besner (1987)

Predictor simulations: non-words

- [Ramscar]Does the frequency of non-words help predict naming latencies?[/Ramscar]
- Re-analysis of naming latencies for non-words in McCann & Besner (1987)
- Frequency is the strongest predictor for non-word naming latencies

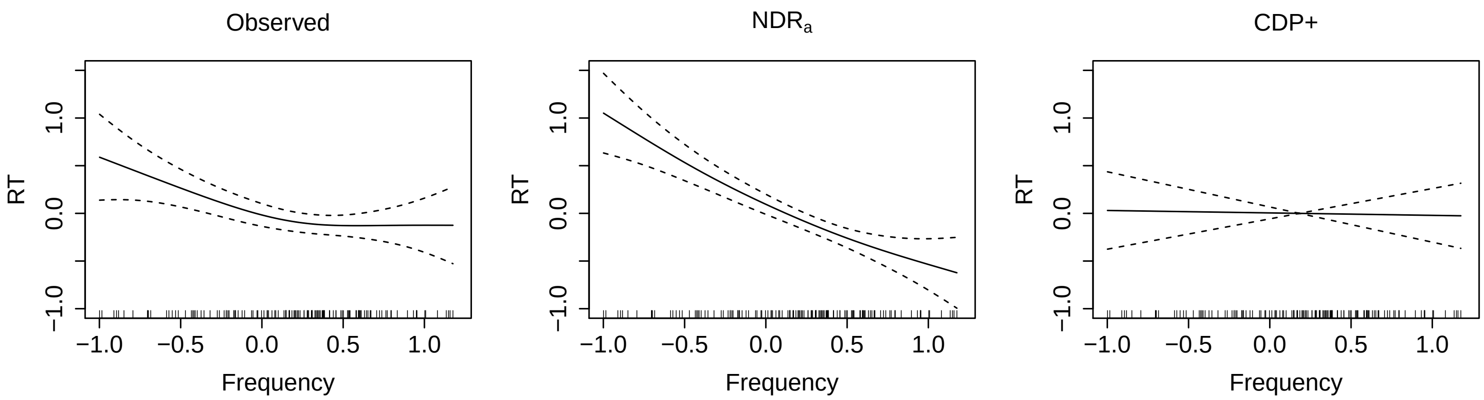
Predictor simulations: non-words

- Difference between words and non-words is graded rather than absolute
- Both words and non-words may or may not have a lexical representation in the mental lexicon of an individual language user
- The probability of a lexical representation is a function of the frequency of a word or non-word
- Fits well with the single-route architecture of the NDRa

Predictor simulations: non-words

- Simulation: retrain NDRa with Google frequencies for non-words
- Does the NDRa capture the effect of non-word frequency?

Predictor simulations: non-words



Outline

- Introduction
- NDRa model
- Simulations
  - Overall model fit
  - Predictor simulations
  - Comparison to dual-route architecture
  - Pronunciation performance
- Conclusions

Comparison to dual-route architecture

- The single-route NDRa model explains a wide range of experimental effects in both word and non-word naming
- Would a sub-lexical route further improve the performance of the NDRa?

Comparison to dual-route architecture

- Add a sub-lexical route to the NDRa
- Discriminative learning network from orthography to phonology
- Does this network help explain additional variance in the observed data?

Comparison to dual-route architecture

	NDR_a	NDR_a^2
Lexical route		
ActLexeme	6.962	5.269
$ActPhon_1$	4.659	4.922
$ActPhon_2$	9.907	9.812
H	−6.178	−6.126
Complexity	18.030	17.266
Non-lexical route		
$ActPhonSub_1$	NA	1.107
$ActPhonSub_1$	NA	−0.525
HSub	NA	−1.942

Comparison to dual-route architecture

- Components of the sub-lexical route do not help explain additional variance
- Correlation with observed naming latencies remains the same
- Conclusion: the addition of a sub-lexical route does not improve the performance of the NDRa model

Outline

- Introduction
- NDRa model
- Simulations
    - Overall model fit
    - Predictor simulations
    - Comparison to dual-route architecture
    - Pronunciation performance
- Conclusions

Pronunciation performance

- Naming latencies reflect bottom-up processes
- Discrimination learning captures bottom-up processing

Pronunciation performance

- Response selection involves top-down processes
- The pre-frontal cortex plays an important role in response conflict resolution
- Functional architecture?

Pronunciation performance

- Provisionary checking mechanism
- Filter set of lexemes that activate demi-syllables based on orthographic overlap with the target word

Pronunciation performance

- Word naming performance: 99.09% (CDP+: 98.68%)
- Nonword naming performance: 70.36% (CDP+: 66.94%)
- With lenient scoring criterion: 98.88% (CDP+: 88.27%)
- Better pronunciation for both words and non-words than CDP+

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Outline

- Introduction
- NDRa model
- Simulations
    - Overall model fit
    - Predictor simulations
    - Comparison to dual-route architecture
- Conclusions

Conclusions

- Discriminative learning works for reading aloud
- A single lexical route is sufficient to explain a wide range of experimental effects in both word and non-word naming

Conclusions

- Outstanding issues:
  - What is the functional architecture of the selection mechanism?
  - Discrete representations are abstractions from the neurobiological reality of language processing
  - Extension to multi-syllabic words

Thank you!