English auxiliary realization and the independence of morphology and phonetics

Laurel MacKenzie and Charles Yang
University of Manchester
University of Pennsylvania

NWAV 41
October 27, 2012

The issue

The issue

The architecture of the grammar:

The issue

The architecture of the grammar:

- **Modular**, feed-forward system with distinct levels of representations and processes?

(Chomsky & Halle 1968; Labov 1969)

The issue

The architecture of the grammar:

- **Modular**, feed-forward system with distinct levels of representations and processes?

(Chomsky & Halle 1968; Labov 1969)

- Interactive, **exemplar-based**, all levels interdependent?

(Pierrehumbert 2002)

A test case

A test case

Phonetic lenition

A test case

Phonetic lenition

- Independent of morphology?

A test case

Phonetic lenition

- Independent of morphology?
- Contingent on “probabilistic relations between words”? (Jurafsky et al. 2001)

Our contribution

Our contribution

Quantitative corpus study of phonetic **h-deletion** in function words

Our contribution

Quantitative corpus study of phonetic **h-deletion** in function words

- Another lenition process that affects multiple lexical items

Our contribution

Quantitative corpus study of phonetic **h-deletion** in function words

- Another lenition process that affects multiple lexical items
- A variable phenomenon that has been little studied empirically

Our contribution

Quantitative corpus study of phonetic **h-deletion** in function words

- Another lenition process that affects multiple lexical items
- A variable phenomenon that has been little studied empirically

Finding: no morphological sensitivity.

Setting the stage: Auxiliary variation

Setting the stage: Auxiliary variation

English auxiliaries vary in phonological shape ("**contraction**")

Setting the stage: Auxiliary variation

English auxiliaries vary in phonological shape ("**contraction**")

Today's talk:

had

has

Setting the stage: Auxiliary variation

3 phonological shapes

Setting the stage: Auxiliary variation

3 phonological shapes

had

has

Setting the stage: Auxiliary variation

3 phonological shapes

	CVC
had	[həd]
has	[həz]

Setting the stage: Auxiliary variation

3 phonological shapes

	CVC	VC
had	[həd]	[əd]
has	[həz]	[əz]

Setting the stage: Auxiliary variation

3 phonological shapes

	CVC	VC	C
had	[həd]	[əd]	[d]
has	[həz]	[əz]	[z]/[s]

Setting the stage: Auxiliary variation

3 phonological shapes,

contingent on host type

	CVC	VC	C
had	[həd]	[əd]	[d]
has	[həz]	[əz]	[z]/[s]

Setting the stage: Auxiliary variation

3 phonological shapes,

contingent on host type

	CVC	VC	C
pron + *had*	[həd]	[əd]	[d]
has	[həz]	[əz]	[z]/[s]

Setting the stage: Auxiliary variation

3 phonological shapes,
contingent on host type

	CVC	VC	C
pron + *had*	[həd]	[əd]	[d]
NP + *has*	[həz]	[əz]	[z]/[s]

Setting the stage: Auxiliary variation

3 phonological shapes,

different grammatical sources

	CVC	VC	C
pron + *had*	[həd]	[əd]	[d]
NP + *has*	[həz]	[əz]	[z]/[s]

(MacKenzie in press)

Setting the stage: Auxiliary variation

3 phonological shapes,

different grammatical sources

	CVC	VC	C
pron + *had*	[həd]	[əd]	[d]
NP + *has*	[həz]	[əz]	[z]/[s]

(MacKenzie in press)

Setting the stage: Auxiliary variation

3 phonological shapes,

different grammatical sources

	CVC	VC	C
pron + *had*	[həd]	[əd]	[d]
NP + *has*	[həz]	[əz]	[z]/[s]

(MacKenzie in press)

Setting the stage: Auxiliary variation

Analysis:

1. morphological

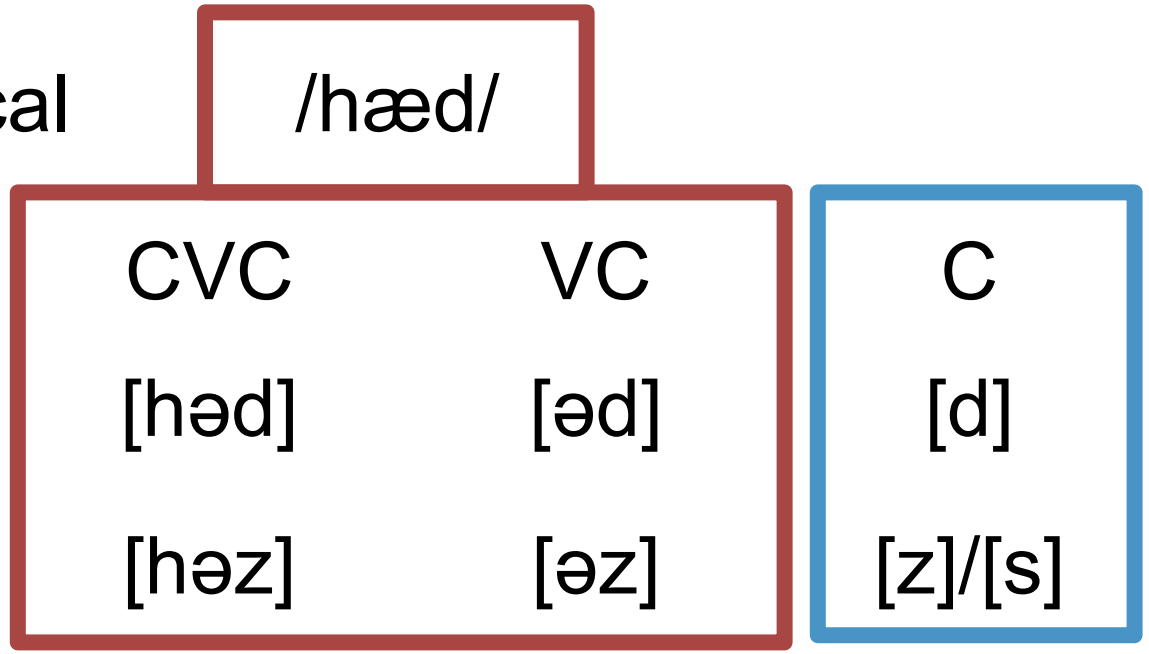
CVC	VC	C
[həd]	[əd]	[d]
[həz]	[əz]	[z]/[s]

(MacKenzie in press)

Setting the stage: Auxiliary variation

Analysis:

1. morphological

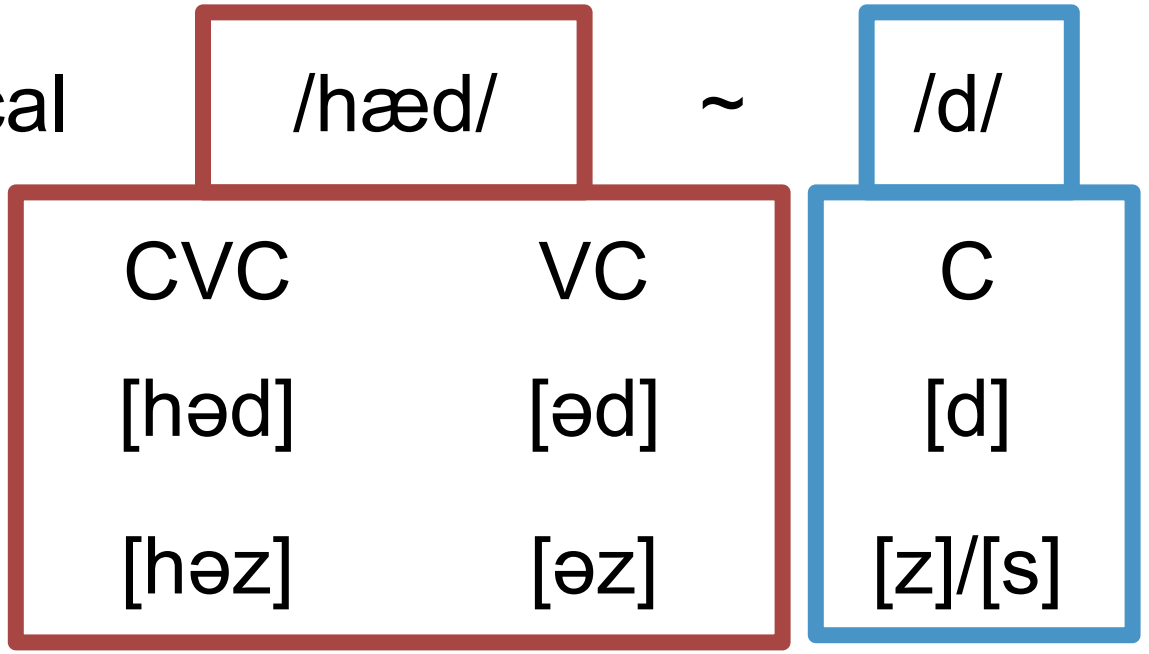


(MacKenzie in press)

Setting the stage: Auxiliary variation

Analysis:

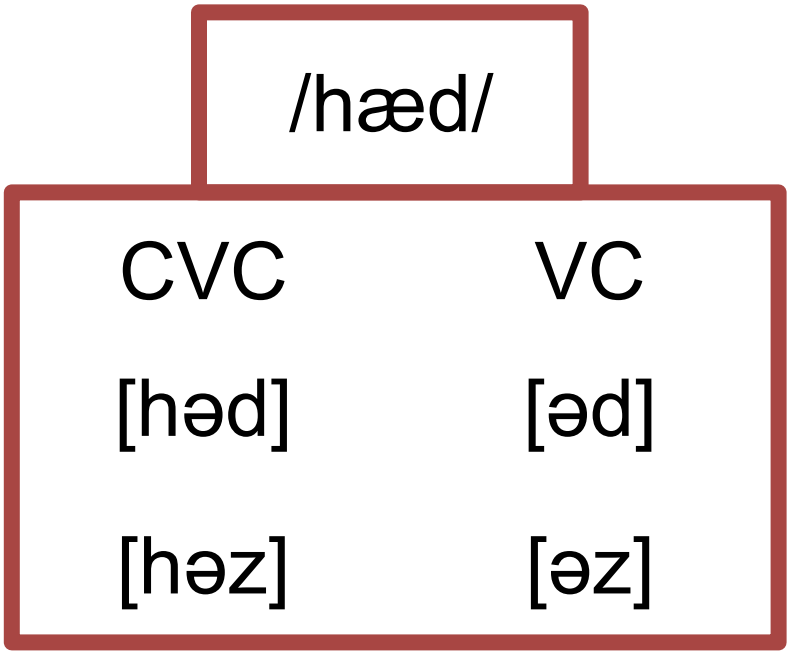
1. morphological



(MacKenzie in press)

Setting the stage: Auxiliary variation

Analysis:



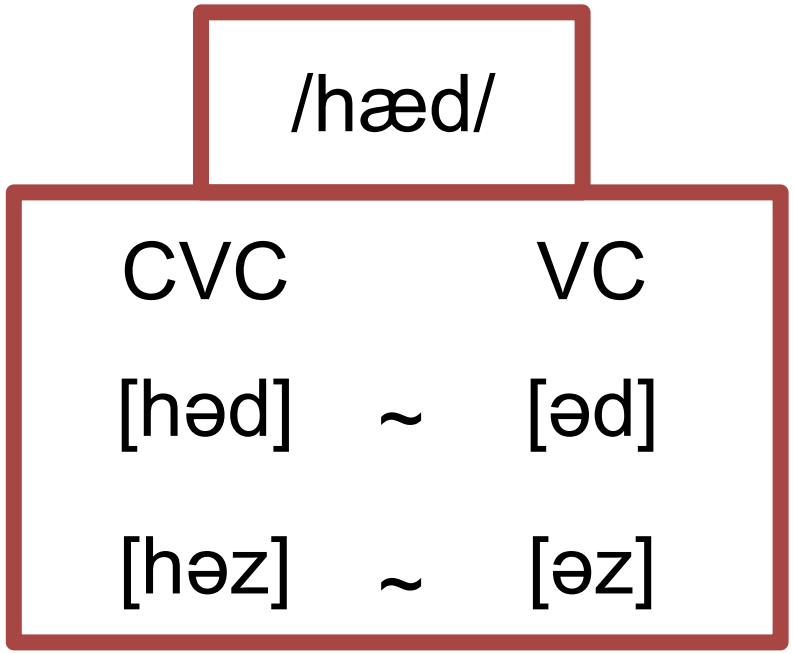
(MacKenzie in press)

Setting the stage: Auxiliary variation

Analysis:

2. phonetic

   h-deletion



(MacKenzie in press)

Setting the stage: Auxiliary variation

Analysis:

2. phonetic

h-deletion

/hæd/

CVC		VC
[həd]	~	[əd]
[həz]	~	[əz]

(MacKenzie in press)

Setting the stage: Auxiliary variation

Analysis:

2. phonetic

h-deletion

/hæd/

CVC		VC
[həd]	~	[əd]
[həz]	~	[əz]

(MacKenzie in press)

Research question

Research question

Is **h-deletion**:

Research question

Is **h-deletion**:

- Independent of morphology?

Research question

Is **h-deletion**:

- Independent of morphology?
- Contingent on “probabilistic relations between words”? (Jurafsky et al. 2001)

Methodology

Corpus: Switchboard (Godfrey et al., 1992)

Methodology

Corpus: Switchboard (Godfrey et al., 1992)

Lexical items:

had

has

Methodology

Corpus: Switchboard (Godfrey et al., 1992)

Lexical items:		/h/	Ø
had	N=284	[həd]	[əd]
has	N=252	[həz]	[əz]

Methodology

Corpus: Switchboard (Godfrey et al., 1992)

Lexical items:		/h/	Ø
had	N=284	[həd]	[əd]
has	N=252	[həz]	[əz]
he	N=273		

Methodology

Corpus: Switchboard (Godfrey et al., 1992)

Lexical items:		/h/	Ø
had	N=284	[həd]	[əd]
has	N=252	[həz]	[əz]
he	N=273		

-conjunctions

Methodology

Corpus: Switchboard (Godfrey et al., 1992)

Lexical items:		/h/	Ø
had	N=284	[həd]	[əd]
has	N=252	[həz]	[əz]
he	N=273		
		-conjunctions	-*wh*-words

Methodology

Corpus: Switchboard (Godfrey et al., 1992)

Lexical items:		/h/	Ø
had	N=284	[həd]	[əd]
has	N=252	[həz]	[əz]
he	N=273		

-conjunctions -*wh*-words

-comp. *that*

Methodology

Corpus: Switchboard (Godfrey et al., 1992)

Lexical items:		/h/	Ø
had	N=284	[həd]	[əd]
has	N=252	[həz]	[əz]
he	N=273		

-conjunctions -*wh*-words

-comp. *that* -*then*

Methodology

Corpus: Switchboard (Godfrey et al., 1992)

Lexical items:		/h/	Ø
had	N=284	[həd]	[əd]
has	N=252	[həz]	[əz]
he	N=273	[hi]	[i]

Methodology

Corpus: Switchboard (Godfrey et al., 1992)

Lexical items:		/h/	Ø
had	N=284	[həd]	[əd]
has	N=252	[həz]	[əz]
he	N=273	[hi]	[i]

Mixed-effects modeling to obtain fitted rates

Findings

Findings

		/h/	Ø	/h/-del.
had	N=284	[həd]	[əd]	**0.38**
has	N=252	[həz]	[əz]	**0.36**
he	N=273	[hi]	[i]	**0.33**

Findings

		/h/	Ø	/h/-del.
had	N=284	[həd]	[əd]	**0.38**
has	N=252	[həz]	[əz]	**0.36**
he	N=273	[hi]	[i]	**0.33**

No effect of lexical item

Interpretation

Interpretation

Consistency of h-deletion across lexical items

	/h/-del.
had	**0.38**
has	**0.36**
he	**0.33**

Interpretation

Consistency of h-deletion across lexical items

- Attributable to some other factor?

	/h/-del.
had	**0.38**
has	**0.36**
he	**0.33**

Interpretation

Consistency of h-deletion across lexical items

- Attributable to some other factor?
  - Item frequency
  - Item predictability

	/h/-del.
had	**0.38**
has	**0.36**
he	**0.33**

(Bybee 2002)

(Jurafsky et al. 2001)

Interpretation

Consistency of h-deletion across lexical items

- Attributable to some other factor?
  - Item frequency
  - Item predictability

calculated in Switchboard

	/h/-del.
had	**0.38**
has	**0.36**
he	**0.33**

(Bybee 2002)

(Jurafsky et al. 2001)

Item frequency

	freq.	/h/-del.
had	18366	**0.38**
has	7640	**0.36**
he	21854	**0.33**

Item frequency

	freq.	/h/-del.
had	18366	**0.38**
has	7640	**0.36**
he	21854	**0.33**

✗

Predictability

	freq.	/h/-del.
had	18366	**0.38**
has	7640	**0.36**
he	21854	**0.33**
	✗	

Predictability

= mean probability given previous word

	freq.	/h/-del.
had	18366	**0.38**
has	7640	**0.36**
he	21854	**0.33**
	✗	

Predictability

= mean probability given previous word

e.g. given ***she***, how likely is ***had*** to follow?

	freq.	/h/-del.
had	18366	**0.38**
has	7640	**0.36**
he	21854	**0.33**
	✗	

Predictability

= mean probability given previous word

e.g. given ***she***, how likely is ***had*** to follow?

	pred.	freq.	/h/-del.
had	0.393	18366	**0.38**
has	0.650	7640	**0.36**
he	0.115	21854	**0.33**
		✗	

Predictability

= mean probability given previous word

e.g. given ***she***, how likely is ***had*** to follow?

	pred.	freq.	/h/-del.
had	0.393	18366	**0.38**
has	0.650	7640	**0.36**
he	0.115	21854	**0.33**
	✗	✗	

Predictability - POS

= mean probability given previous part of speech

	pred.	freq.	/h/-del.
had	0.393	18366	**0.38**
has	0.650	7640	**0.36**
he	0.115	21854	**0.33**
	✗	✗	

Predictability - POS

= mean probability given previous part of speech

e.g. given ***pro***, how likely is ***had*** to follow?

	pred.	freq.	/h/-del.
had	0.393	18366	**0.38**
has	0.650	7640	**0.36**
he	0.115	21854	**0.33**
	✗	✗	

Predictability - POS

= mean probability given previous part of speech

e.g. given ***pro***, how likely is ***had*** to follow?

	POS-pred.	pred.	freq.	/h/-del.
had	0.063	0.393	18366	**0.38**
has	0.005	0.650	7640	**0.36**
he	0.014	0.115	21854	**0.33**
	✗	✗	✗	

Collocation with following word

Collocation with following word

Each item's most frequent successor:

Collocation with following word

Each item's most frequent successor:

had ***been***

Collocation with following word

Each item's most frequent successor:

had ***been***

has ***been***

Collocation with following word

Each item's most frequent successor:

had ***been***

has ***been***

he ***had***

Collocation with following word

Each item's most frequent successor:

	/h/-del.
had ***been***	**0.37**
has ***been***	**0.28**
he ***had***	**0.29**

Collocation with following word

All other successors:

Collocation with following word

All other successors:

had ¬***been***

Collocation with following word

All other successors:

had ¬***been***

has ¬***been***

Collocation with following word

All other successors:

had ¬***been***

has ¬***been***

he ¬***had***

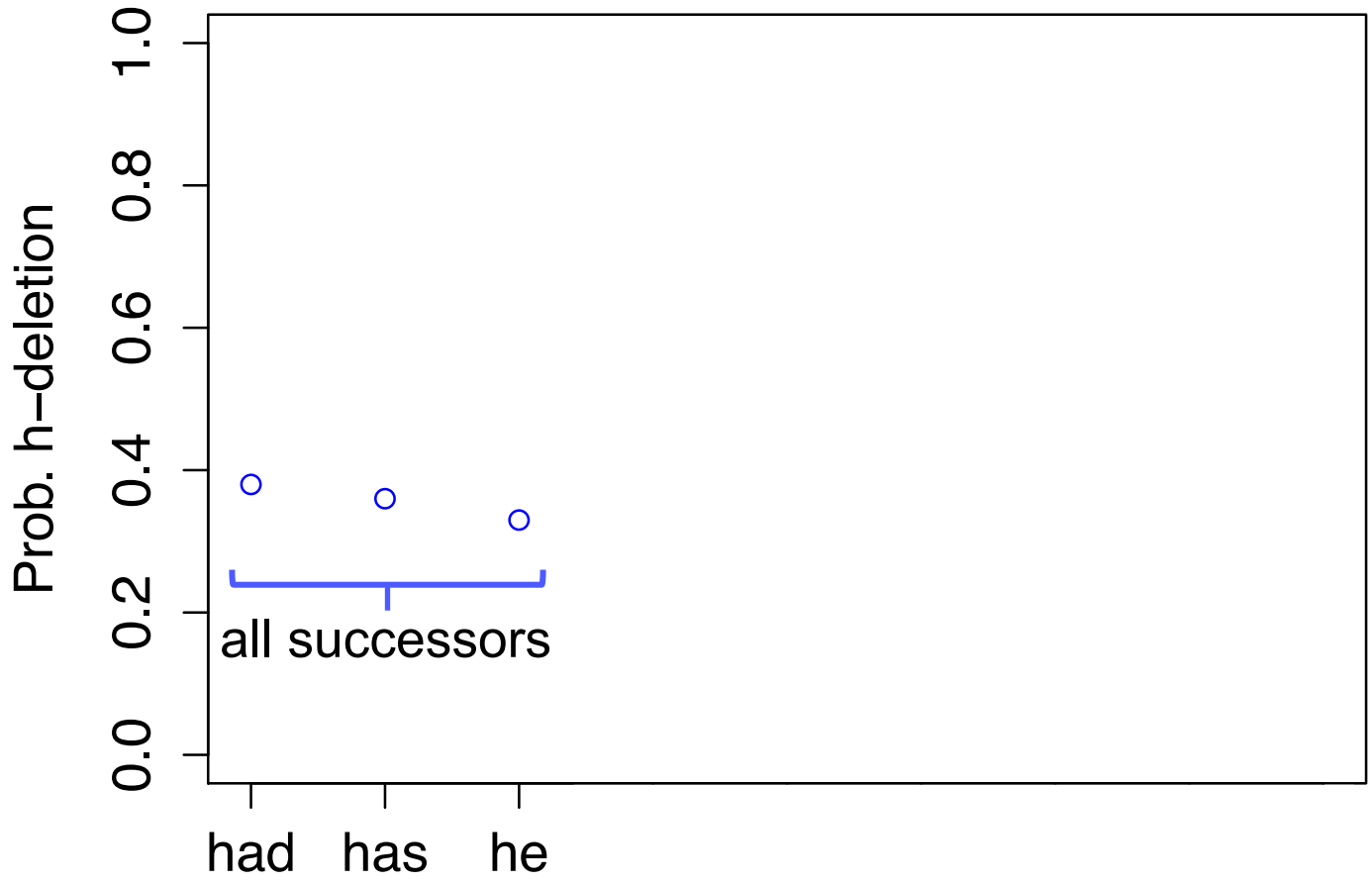
Collocation with following word

All other successors:

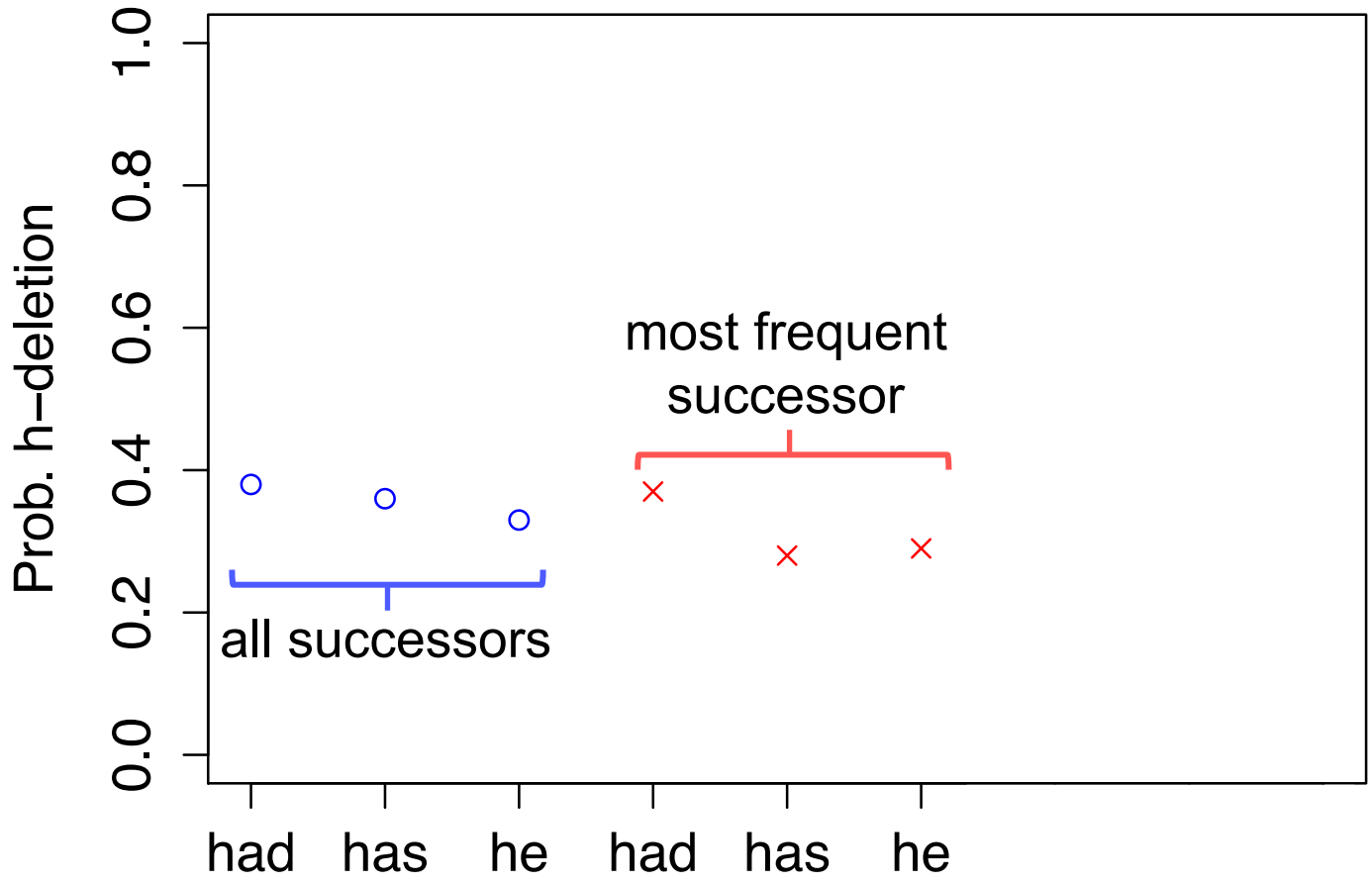
	/h/-del.
had ¬***been***	**0.33**
has ¬***been***	**0.41**
he ¬***had***	**0.39**

h-deletion by successor

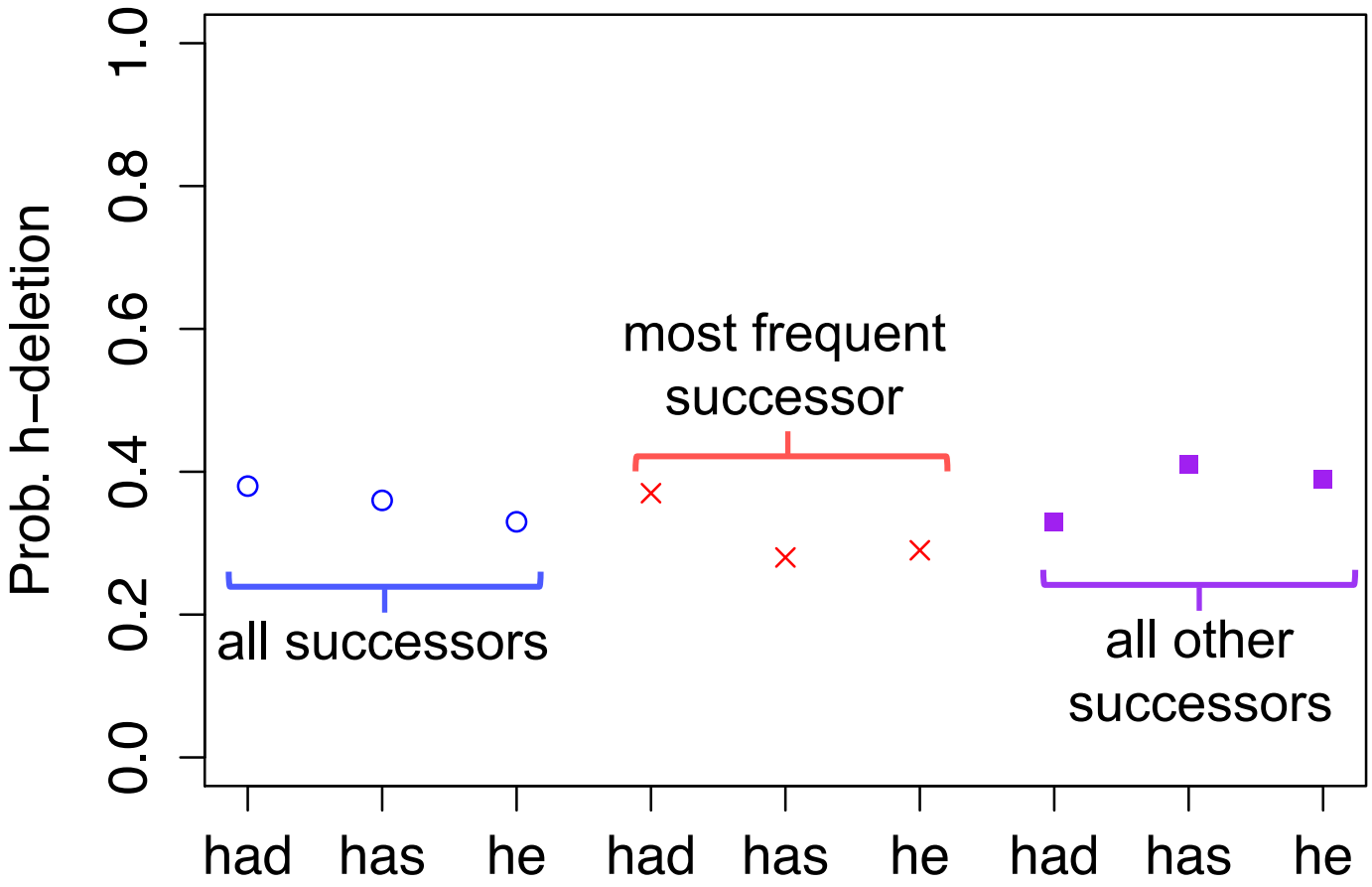
h-deletion by successor



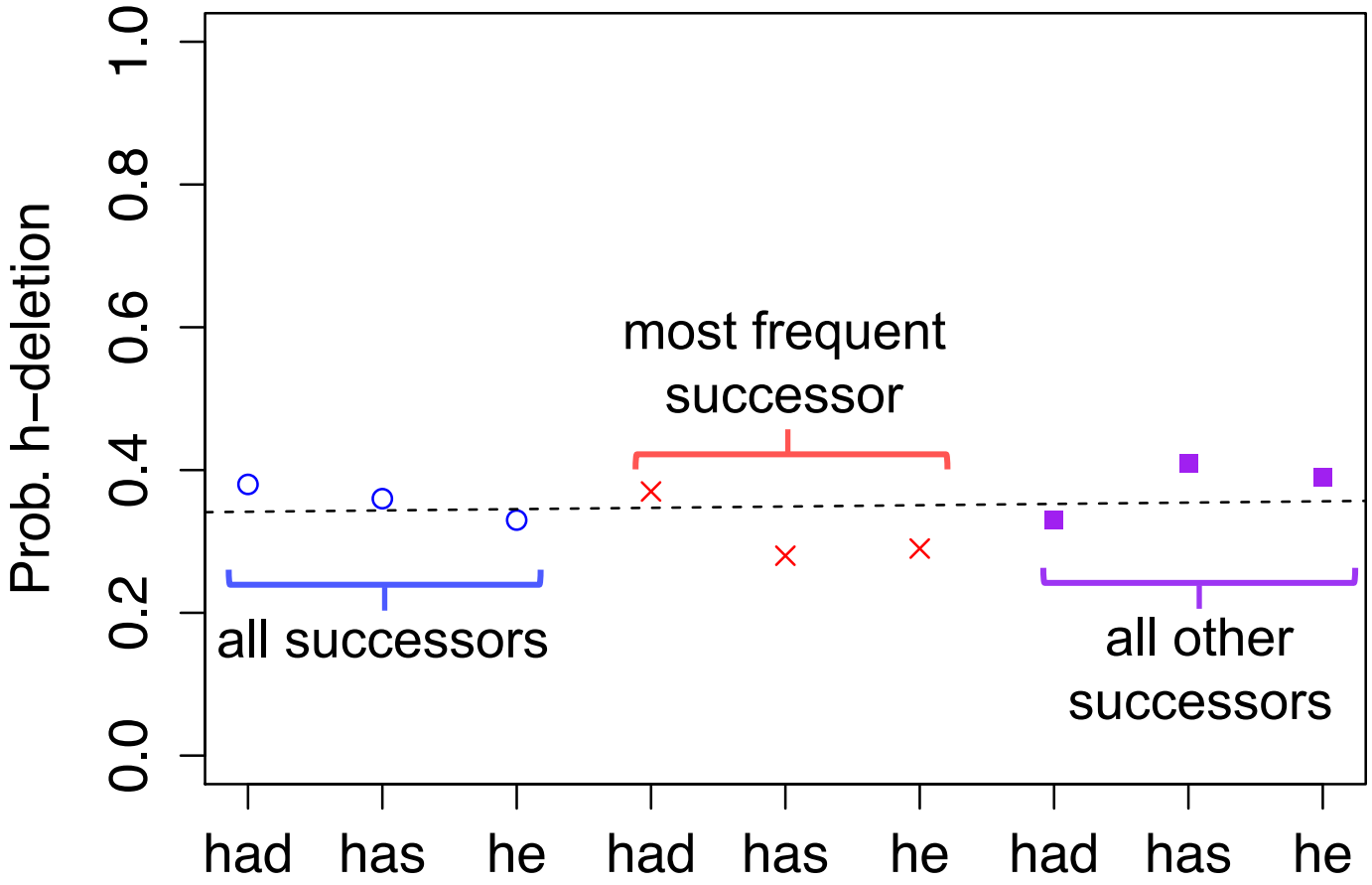
h-deletion by successor



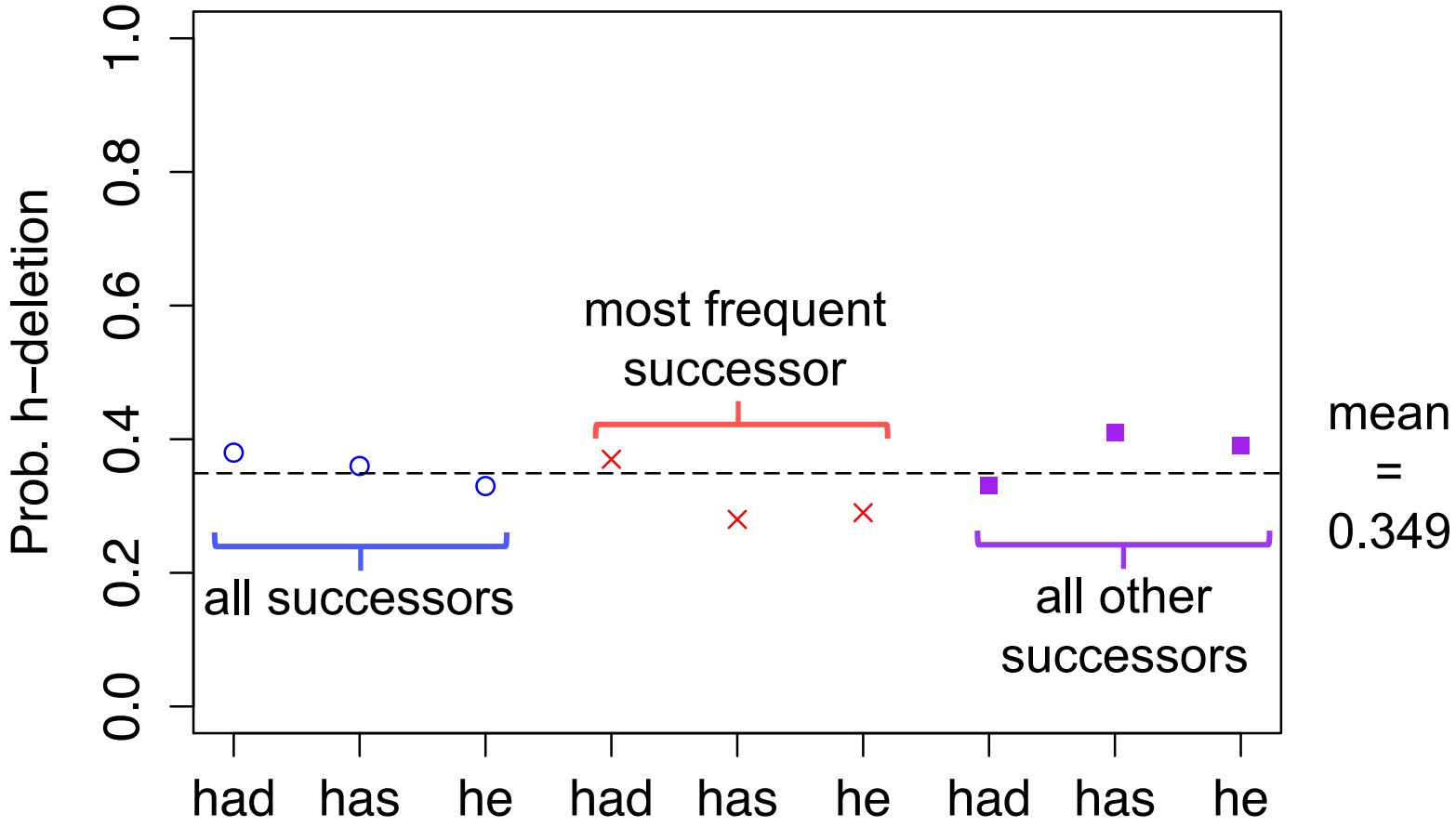
h-deletion by successor



h-deletion by successor



h-deletion by successor



In sum

In sum

h-deletion applies at a consistent rate, irrespective of

In sum

h-deletion applies at a consistent rate, irrespective of

- Item frequency

In sum

h-deletion applies at a consistent rate, irrespective of

- Item frequency
- Item predictability given preceding word / preceding part of speech

In sum

h-deletion applies at a consistent rate, irrespective of

- Item frequency
- Item predictability given preceding word / preceding part of speech
- Predictability of following word given item

Implications

Implications

- A fast-speech lenition process that makes no reference to morphology

Implications

- A fast-speech lenition process that makes no reference to morphology

- Modular separation of morphology/phonetics

Implications

- A fast-speech lenition process that makes no reference to morphology
- Modular separation of morphology/phonetics
- Challenges for exemplar-based models which predict increased lenition with greater frequency/predictability

Future work

Future work

- More corpora

Future work

- More corpora
  - Phila. corpus: *had* @ 0.37, *has* @ 0.39

Future work

- More corpora
  - Phila. corpus: *had* @ 0.37, *has* @ 0.39
- More items?

Future work

- More corpora
  - Phila. corpus: *had* @ 0.37, *has* @ 0.39
- More items?
  - Main verb *have*?

Future work

- More corpora
  - Phila. corpus: *had* @ 0.37, *has* @ 0.39
- More items?
  - Main verb *have*?
- Connections between /h/-del. & other processes

Future work

- More corpora
  - Phila. corpus: *had* @ 0.37, *has* @ 0.39
- More items?
  - Main verb *have*?
- Connections between /h/-del. & other processes
  - Flapping / glottalization

Future work

- More corpora
  - Phila. corpus: *had* @ 0.37, *has* @ 0.39
- More items?
  - Main verb *have*?
- Connections between /h/-del. & other processes
  - Flapping / glottalization
  - Linking /r/ (Ogden 1999)

Thank you!

- Acoustic measures of /h/ (Pierrehumbert & Talkin 1992)

Raw frequency of *has/had*

Calculated by searching for item followed by participle, with zero or one words intervening (covers negation/adverbs)

Predictability from preceding POS

- *had:* most likely to follow pronouns & nouns
- *has:* most likely to follow pronouns & nouns
- *he:* most likely to follow conjunctions & prepositions

Successor frequency given h-word

- *had:*
  - #1: *been* (16%)
  - #2: *done* (6%)
- *has*:
  - #1: *been* (30%)
  - #2: *gone* (8%)