Before the rise of *um*

Derek Denis and Tim Gadanidis

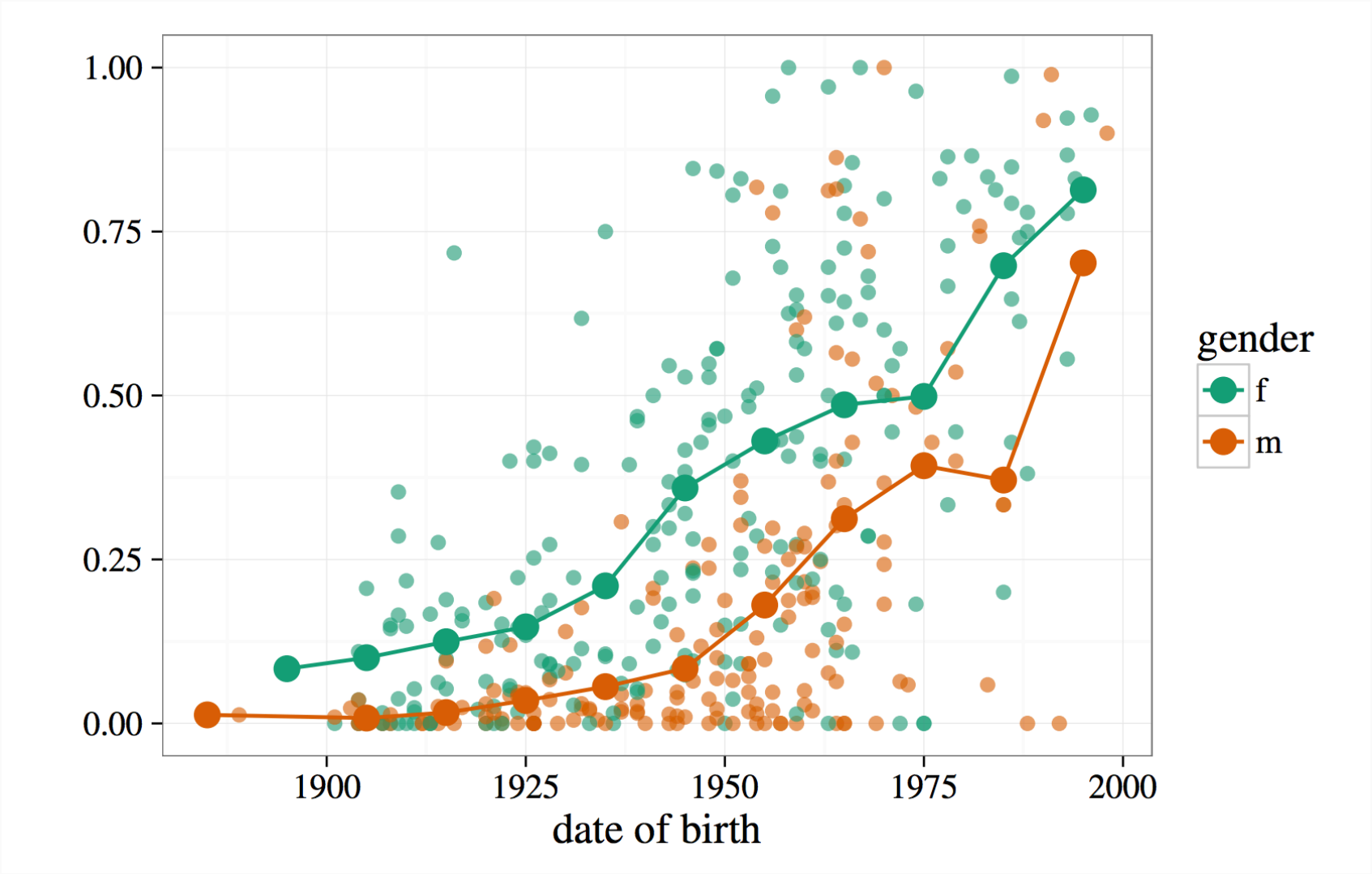
University of Toronto (Mississauga)

DiPVaC4

May 28, 2018

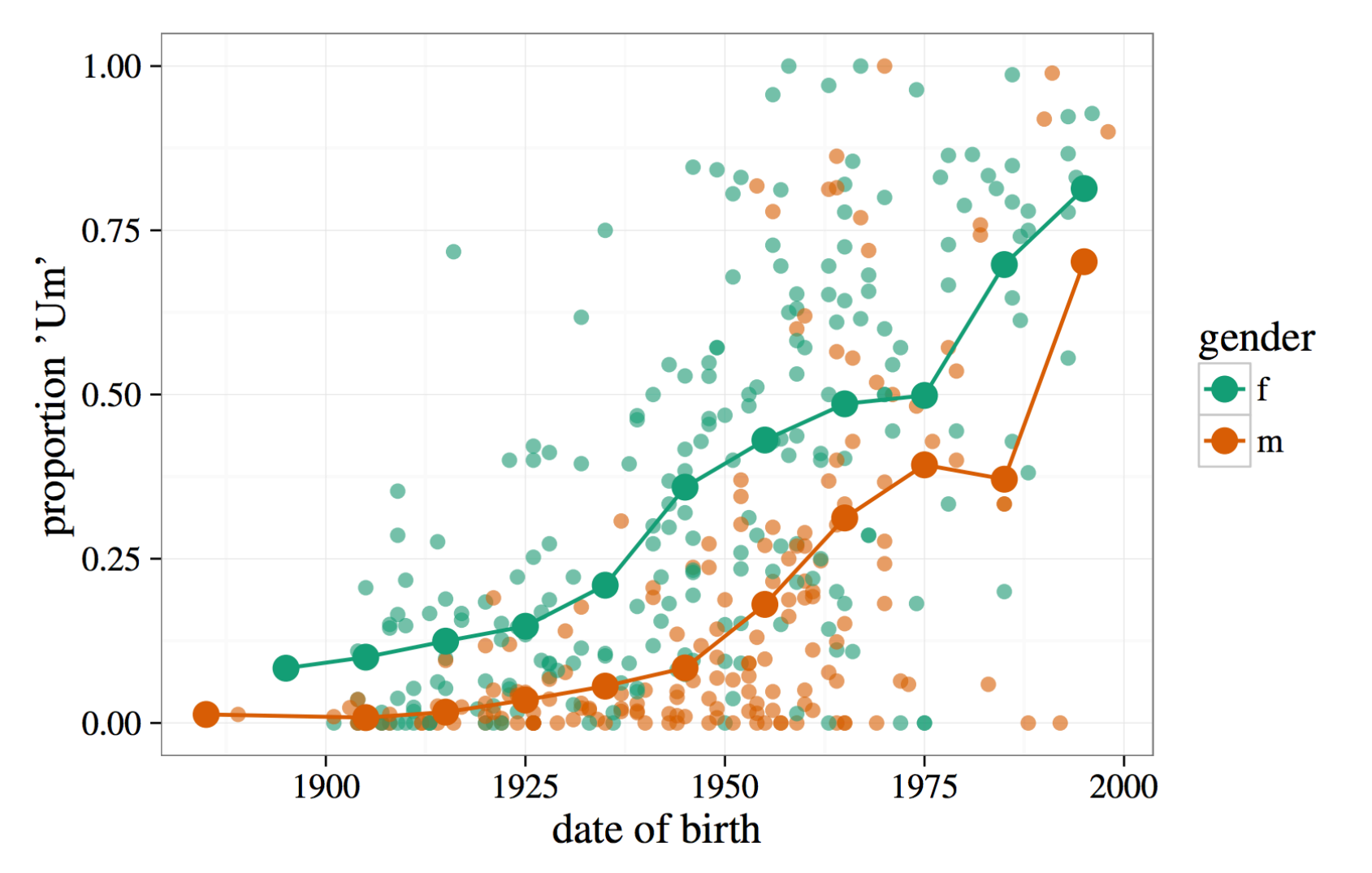
Introduction

“A textbook language change in progress”



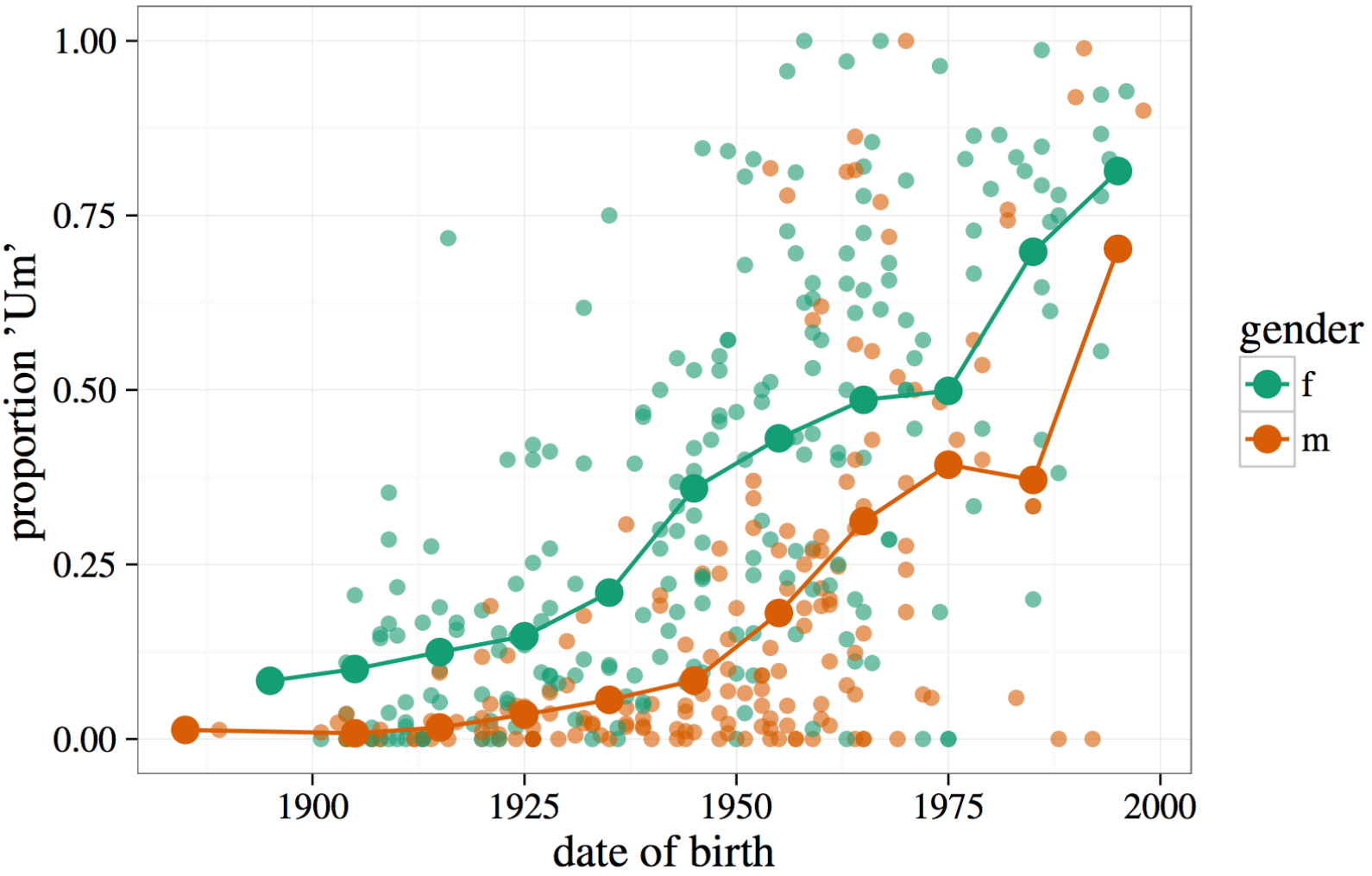
(Fruehwald, 2016: 43)

“A textbook language change in progress”



(Fruehwald, 2016: 43)

“A textbook language change in progress”



(UHM): *“undergoing a language change just like any other” (p. 46)*

(Fruehwald, 2016: 43)

Previous views on (UHM)

Psycholinguistic/processing: ‘filled pauses’/‘disfluencies’

- Levelt (1989): (UHM) is a symptom of processing error

Previous views on (UHM)

Psycholinguistic/processing: 'filled pauses'/'disfluencies'

- Levelt (1989): (UHM) is a symptom of processing error

Discourse/interactional: 'hesitation markers'

- Maclay and Osgood (1959): (UHM) is used for floor/turn management
- Clark and Fox Tree (2002): (UHM) signals an incoming delay in speech; *uh* signals a short delay while *um* signals a long one.

Previous views on (UHM)

Psycholinguistic/processing: 'filled pauses'/'disfluencies'

- Levelt (1989): (UHM) is a symptom of processing error

Discourse/interactional: 'hesitation markers'

- Maclay and Osgood (1959): (UHM) is used for floor/turn management
- Clark and Fox Tree (2002): (UHM) signals an incoming delay in speech; *uh* signals a short delay while *um* signals a long one.

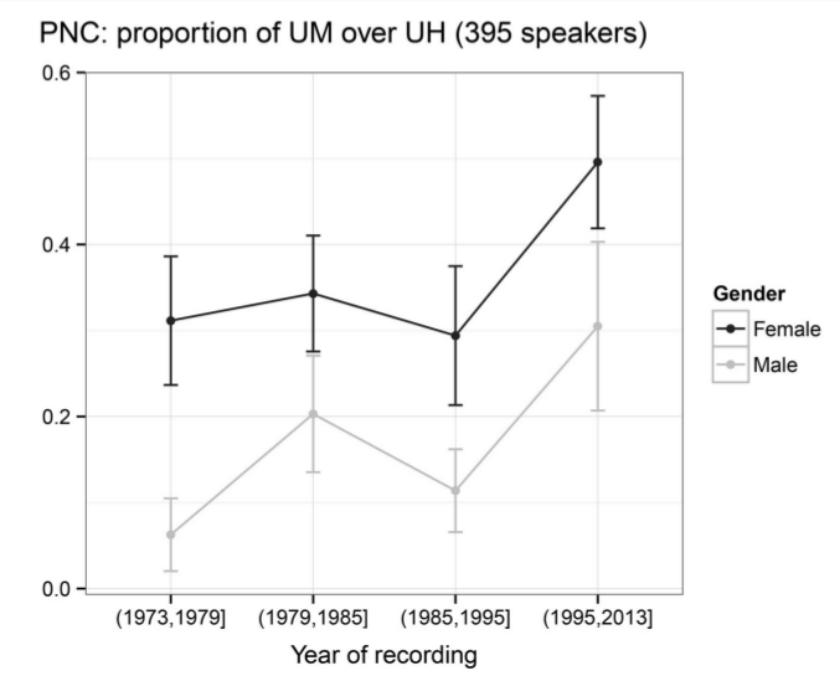
Pragmatic/interpersonal: 'pragmatic marker'

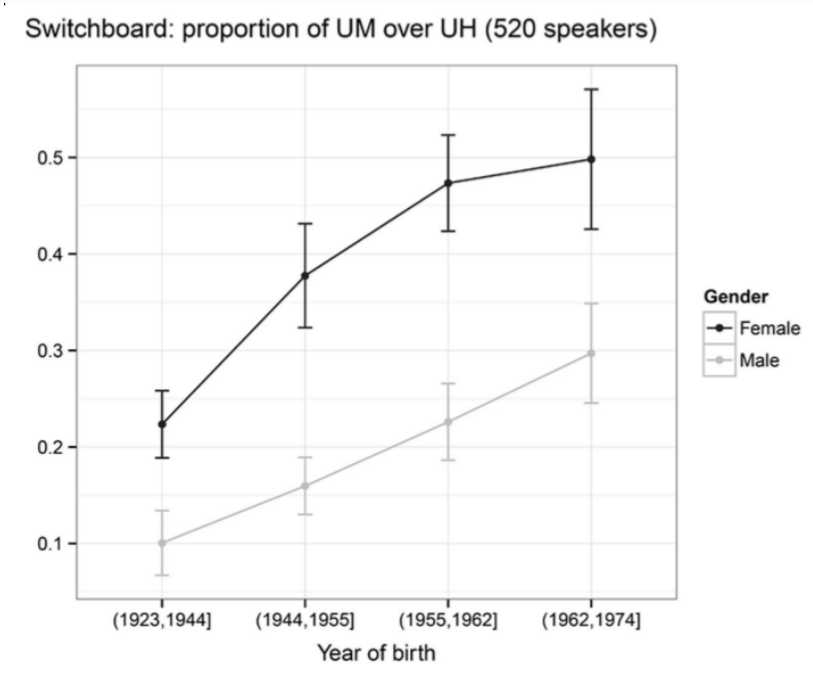
- Tottie (2016): (UHM) is a pragmatic marker indicating speaker planning ('planners'; 'I'm thinking').
- Tottie (2017): (UHM) in writing function as stance adverbs.
  - *Uh*, that's hogwash. (p. 15)
  - A sudden bear made a huge crash in the forest. One guy, *um*, taking care of business, came running back into camp.
- Gadanidis (2018): (UHM) in CMC has a range of sociopragmatic meanings.

Change in progress

There's also a clear change in progress underway in real-time and apparent-time.

- Fruehwald (2016): Women and young people more likely to use *um* than *uh*
- Wieling et al. (2016): Same trend across five other Germanic languages[1]





[1] Dutch, German, Norwegian, Danish, Faroese

Our talk

Our question: why *um*?

- We investigate the early stage of the rise of *um* and ask:

Our talk

Our question: why *um*?

- We investigate the early stage of the rise of *um* and ask:
  - What was the trigger for this “textbook” change?

Our talk

Our question: why *um*?

- We investigate the early stage of the rise of *um* and ask:
  - What was the trigger for this “textbook” change?
    - **Hypothesis**: a new function of (UHM) led to the rise of *um*.
    - We test functional expansion *cum* positional differences: “turn initial *um* may be the best candidate for a new discourse function coming into use” (Fruehwald 2016: 46).

Our talk

Our question: why *um*?

- We investigate the early stage of the rise of *um* and ask:
  - What was the trigger for this "textbook" change?
    - **Hypothesis**: a new function of (UHM) led to the rise of *um*.
    - We test functional expansion *cum* positional differences: "turn initial *um* may be the best candidate for a new discourse function coming into use" (Fruehwald 2016: 46).
- A (modified) variationist approach:
  - First, we consider the relative proportions of *um* and *uh*; a traditional linguistic variable.
  - Second, we consider the relative frequency of the (UHM) variable itself to help interpret the patterns we find with the proportions.

Data, methods

Farm Work and Farm Life Since 1890 Oral Histories

- Oral history interviews with 155 elderly farmers.

Farm Work and Farm Life Since 1890 Oral Histories

- Oral history interviews with 155 elderly farmers.
- Recorded in 1984:
  - Born between 1891 and 1919.

Farm Work and Farm Life Since 1890 Oral Histories

- Oral history interviews with 155 elderly farmers.
- Recorded in 1984:
  - Born between 1891 and 1919.
- Five regions in Ontario, Canada
  - Timiskaming
  - Essex
  - Dufferin
  - Niagara Region
  - Eastern Ontario

Farm Work and Farm Life Since 1890 Oral Histories

- Oral history interviews with 155 elderly farmers.
- Recorded in 1984:
  - Born between 1891 and 1919.
- Five regions in Ontario, Canada
  - Timiskaming
  - Essex
  - Dufferin
  - Niagara Region
  - Eastern Ontario

Farm Work and Farm Life Since 1890 Oral Histories

- Oral history interviews with 155 elderly farmers.
- Recorded in 1984:
  - Born between 1891 and 1919.
- Five regions in Ontario, Canada
  - Timiskaming
  - Essex
  - Dufferin
  - Niagara Region
  - Eastern Ontario
- Oral histories “are a vestige of an earlier vernacular, its structure, its features, its grammar, its sounds” (Denis, 2016).

Extraction and coding

- Extracted every instance of `'uh'` or `'um'` from transcripts
  - Transcription protocol emphasised faithful reproduction of *uh* and *um*
  - Unrelated uses like `'uh oh'` were excluded
- Also extracted data from interviewers, which we analyze separately
- Coded for the following factors:
  - Social factors (year of birth, gender, region)
  - Utterance position (initial or non-initial)
    - e.g. "Uh as a rule they harrowed it before they um drilled it."
    - Including initial collocations with other discourse-pragmatic markers like `'well'`, and cliticized `'and-uh'`, etc.
    - So utterance-initial does not have to be turn-initial (c.f. Fruehwald, 2016; Wieling et al., 2016)
  - Cliticization, e.g. `'and-uh' 'but-uh'` (following Tottie, 2017)

Results

Cross-community comparison

Community	Raw N *uh*	Raw N *um*	% *um*	Mean *uh* /1000	Mean *um*/1000	Mean UHM/1000
Niagara	1864	357	16.1	21.3	4.1	25.4
E. Ont.	1563	168	9.7	22.6	2.4	25.0
F-INT	321	318	49.8	12.4	12.3	24.7
M-INT	255	51	16.7	13.2	2.6	15.8
S. Board	—	—	28.3	22.1	7.5	29.6
Fisher	—	—	64.1	6.8	9.9	16.7
PNC	—	—	27.6	13.2	4.5	17.7
BNC	—	—	46.1	4.5	4.3	8.8

Cross-community comparison

Community	Raw N *uh*	Raw N *um*	% *um*	Mean *uh* /1000	Mean *um*/1000	Mean UHM/1000
Niagara	1864	357	16.1	21.3	4.1	25.4
E. Ont.	1563	168	9.7	22.6	2.4	25.0
F-INT	321	318	49.8	12.4	12.3	24.7
M-INT	255	51	16.7	13.2	2.6	15.8
S. Board	—	—	28.3	22.1	7.5	29.6
Fisher	—	—	64.1	6.8	9.9	16.7
PNC	—	—	27.6	13.2	4.5	17.7
BNC	—	—	46.1	4.5	4.3	8.8

- Our data have low proportion of *um* compared to other corpora.

Cross-community comparison

Community	Raw N *uh*	Raw N *um*	% *um*	Mean *uh* /1000	Mean *um*/1000	Mean UHM/1000
Niagara	1864	357	16.1	21.3	4.1	25.4
E. Ont.	1563	168	9.7	22.6	2.4	25.0
F-INT	321	318	49.8	12.4	12.3	24.7
M-INT	255	51	16.7	13.2	2.6	15.8
S. Board	—	—	28.3	22.1	7.5	29.6
Fisher	—	—	64.1	6.8	9.9	16.7
PNC	—	—	27.6	13.2	4.5	17.7
BNC	—	—	46.1	4.5	4.3	8.8

- Our data have low proportion of *um* compared to other corpora.
  - But female interviewer is substantially more *um*-full

Cross-community comparison

Community	Raw N *uh*	Raw N *um*	% *um*	Mean *uh* /1000	Mean *um*/1000	Mean UHM/1000
Niagara	1864	357	16.1	21.3	4.1	25.4
E. Ont.	1563	168	9.7	22.6	2.4	25.0
F-INT	321	318	49.8	12.4	12.3	24.7
M-INT	255	51	16.7	13.2	2.6	15.8
S. Board	—	—	28.3	22.1	7.5	29.6
Fisher	—	—	64.1	6.8	9.9	16.7
PNC	—	—	27.6	13.2	4.5	17.7
BNC	—	—	46.1	4.5	4.3	8.8

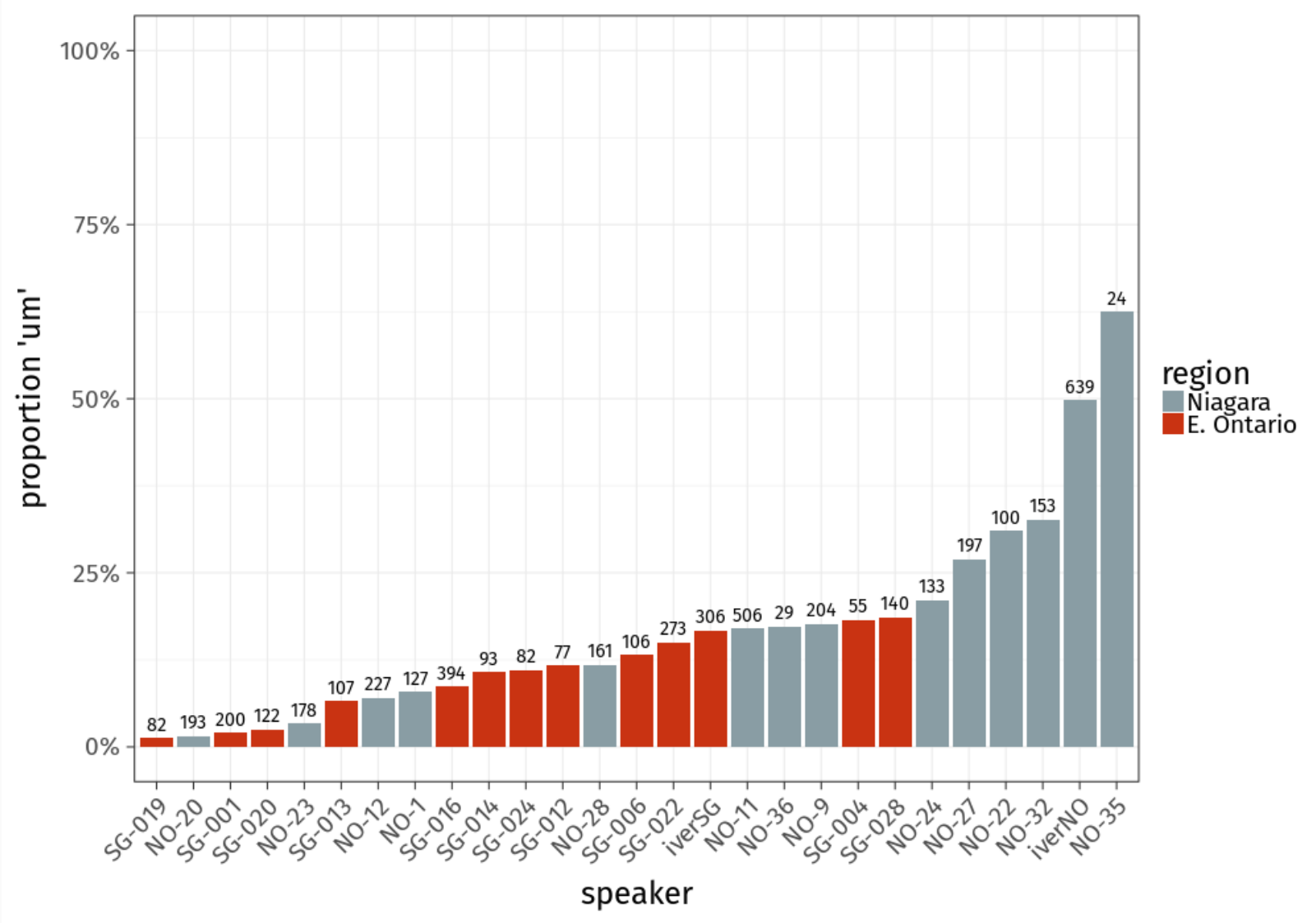
- Our data have low proportion of *um* compared to other corpora.
    - But female interviewer is substantially more *um*-full
- Relative frequency of the UHM variable in our data is on par with other corpora **but** ...

Cross-community comparison

Community	Raw N *uh*	Raw N *um*	% *um*	Mean *uh* /1000	Mean *um*/1000	Mean UHM/1000
Niagara	1864	357	16.1	21.3	4.1	25.4
E. Ont.	1563	168	9.7	22.6	2.4	25.0
F-INT	321	318	49.8	12.4	12.3	24.7
M-INT	255	51	16.7	13.2	2.6	15.8
S. Board	—	—	28.3	22.1	7.5	29.6
Fisher	—	—	64.1	6.8	9.9	16.7
PNC	—	—	27.6	13.2	4.5	17.7
BNC	—	—	46.1	4.5	4.3	8.8

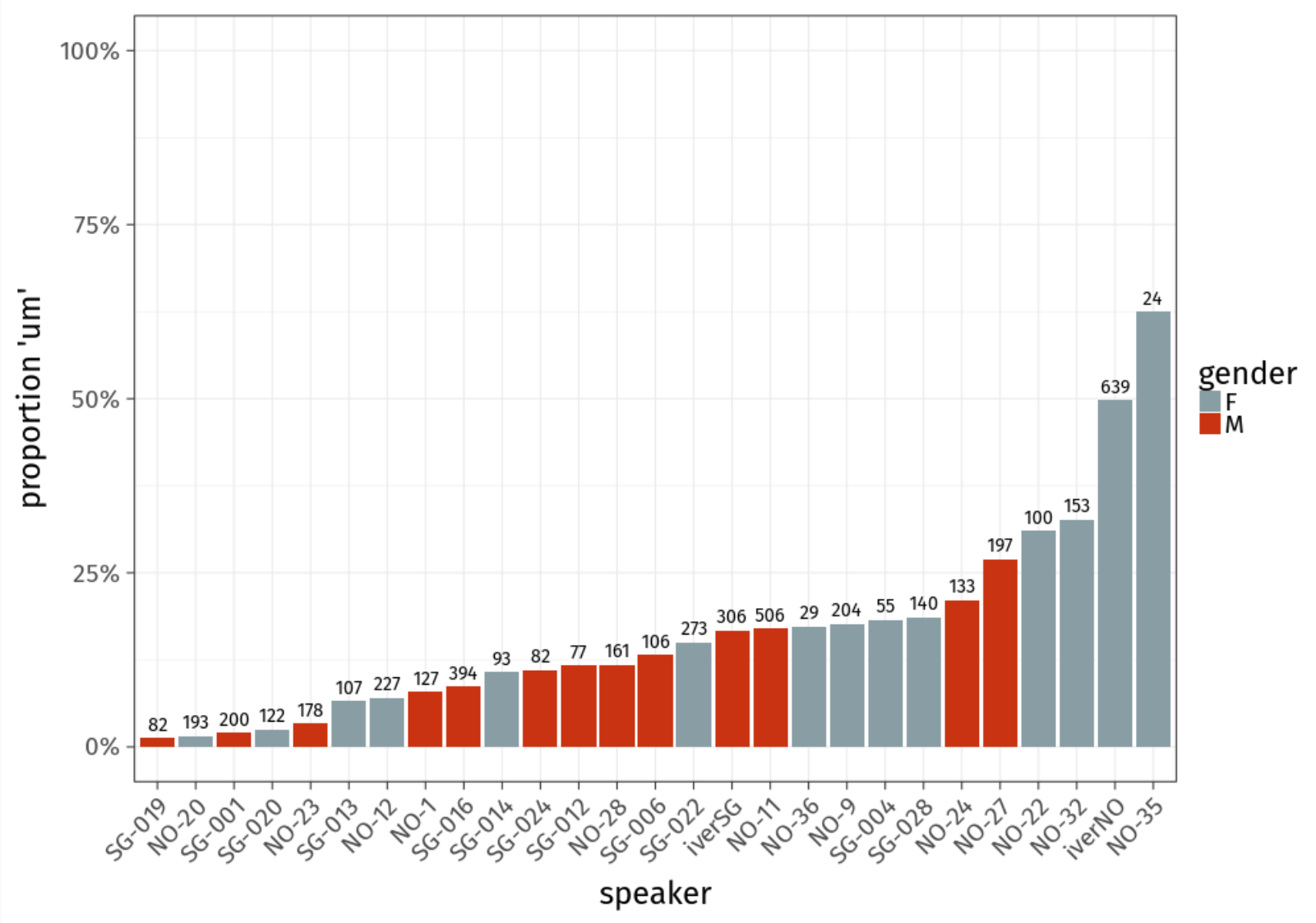
- Our data have low proportion of *um* compared to other corpora.
  - But female interviewer is substantially more *um*-full
- Relative frequency of the UHM variable in our data is on par with other corpora **but** ...
  - We are very cautious about making this comparison due to extremely different methods of data collection and transcription (c.f. Pichler, 2010).

Proportion per speaker by region



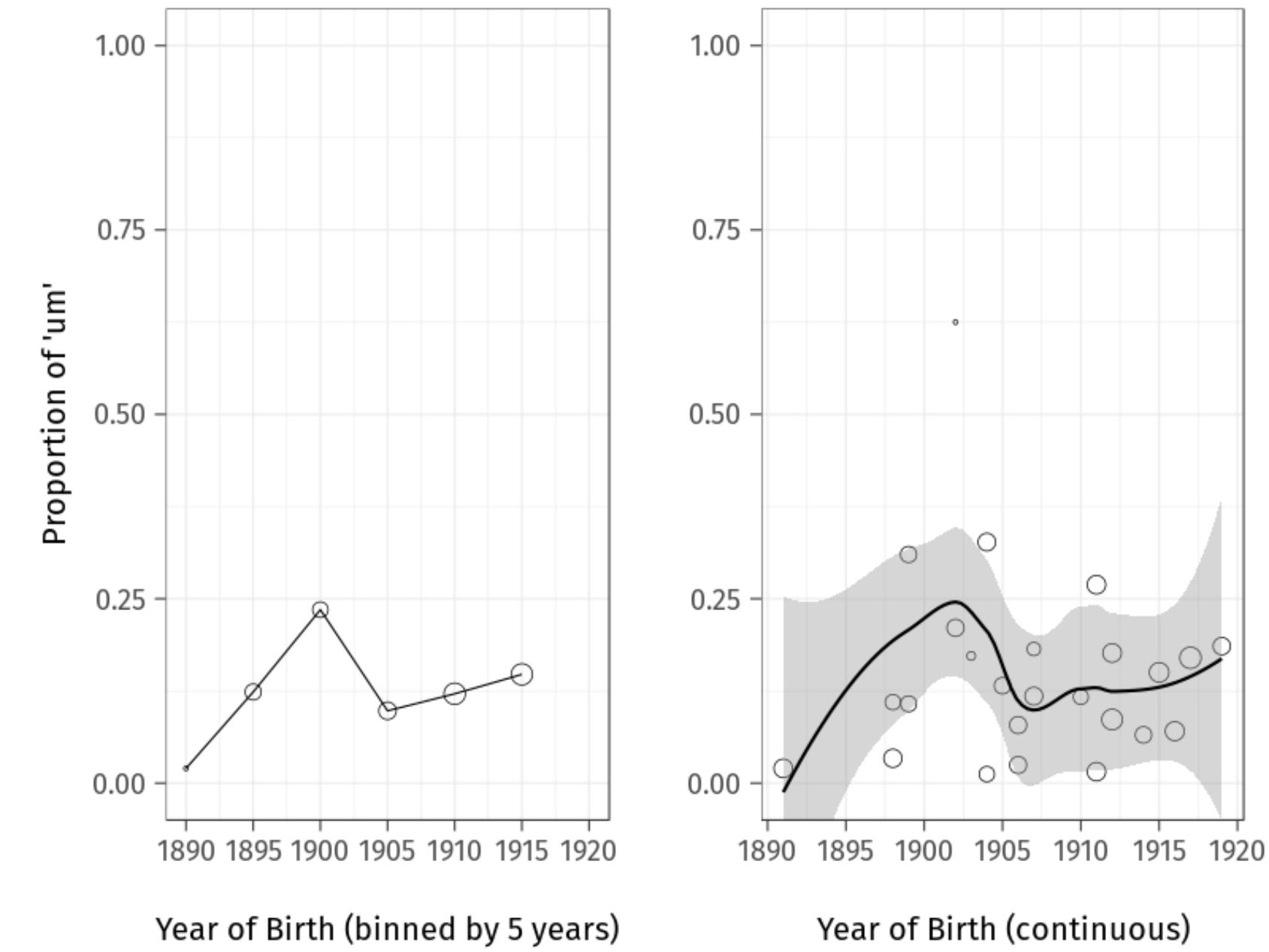
- All speakers are variable; most range between 5-30% *um*
- No clear pattern by region

Proportion per speaker by gender

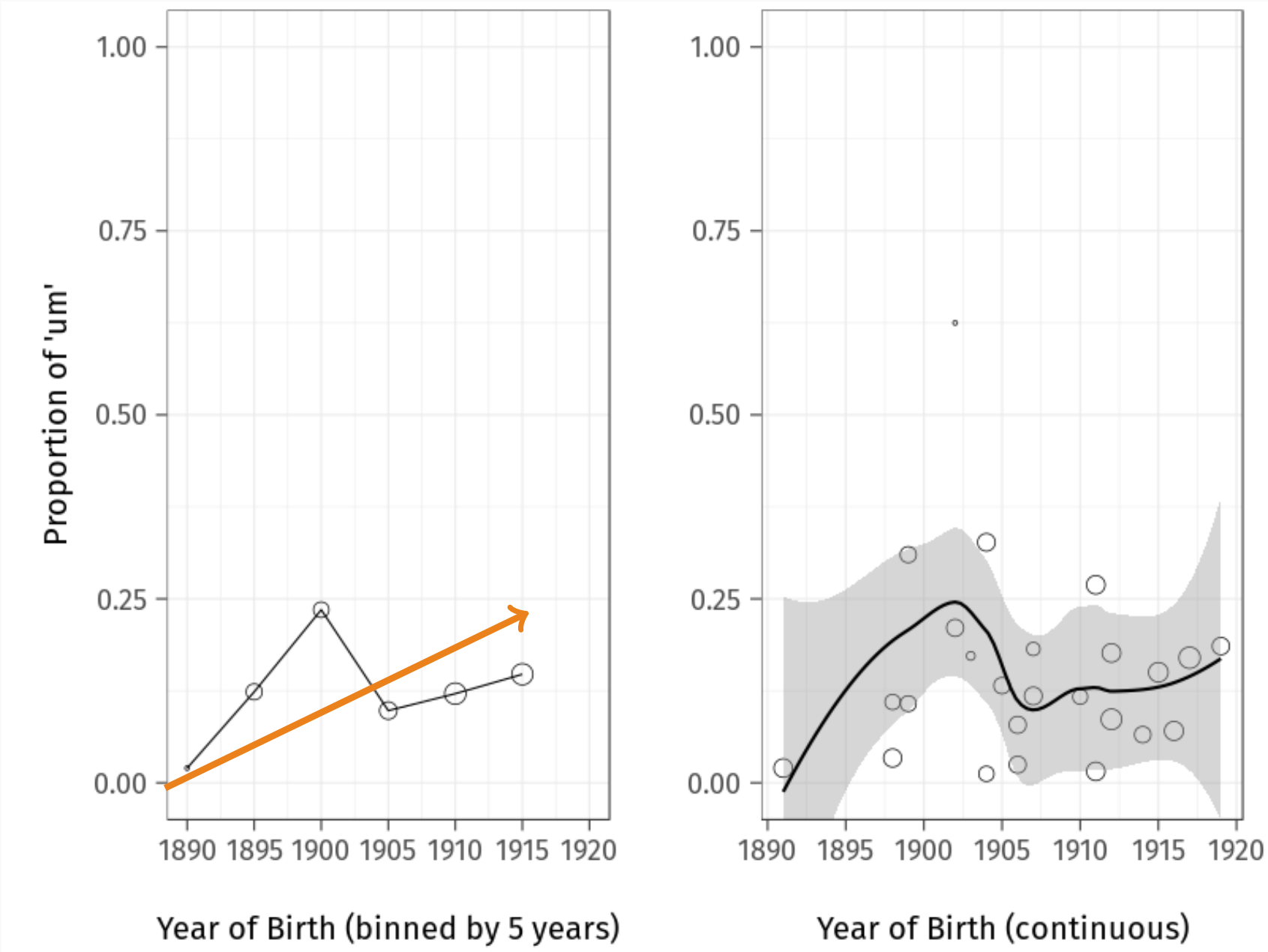


- All speakers are variable; most range between 5-30% *um*
- No clear pattern by gender

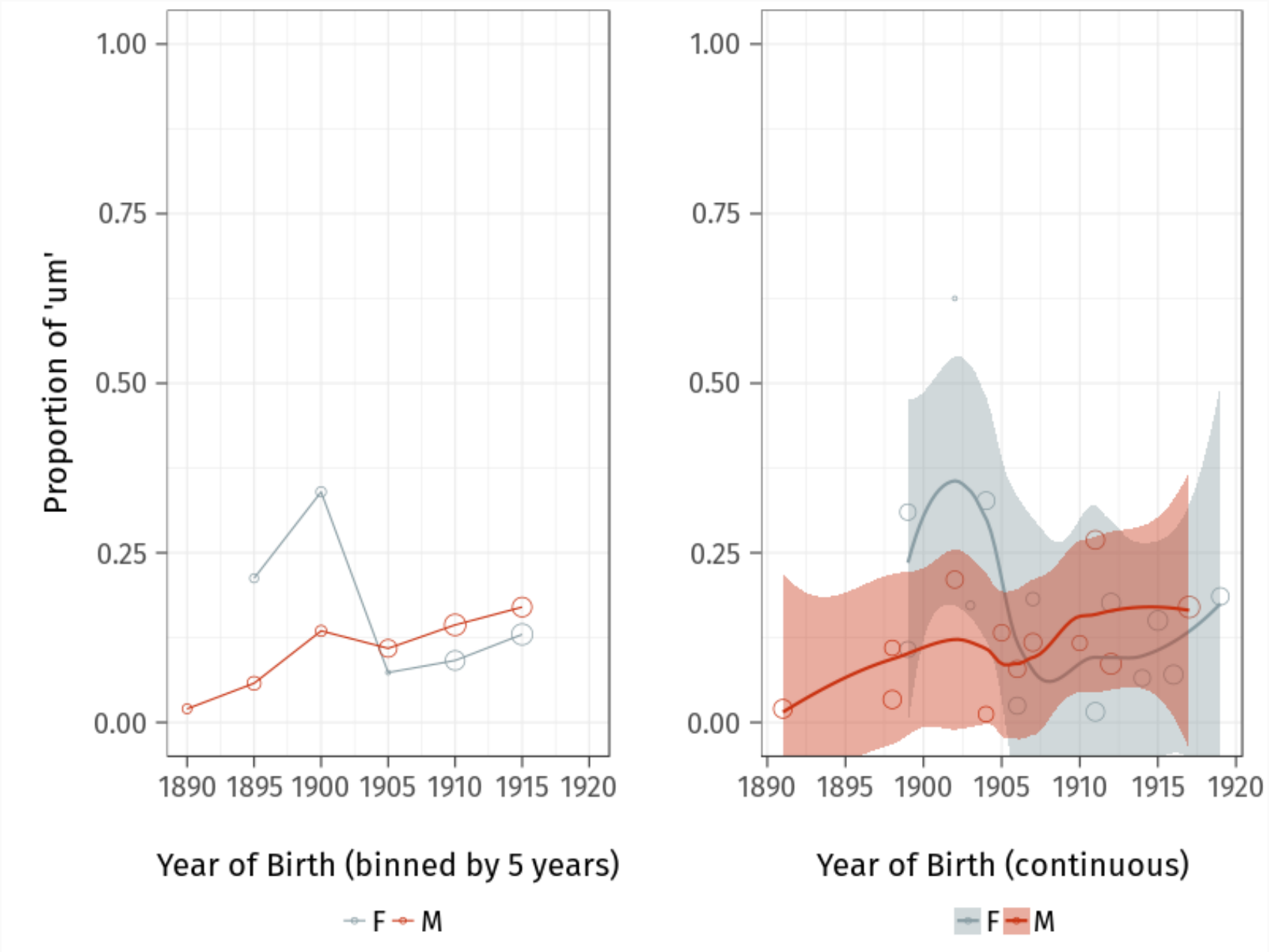
Proportional frequency in apparent time



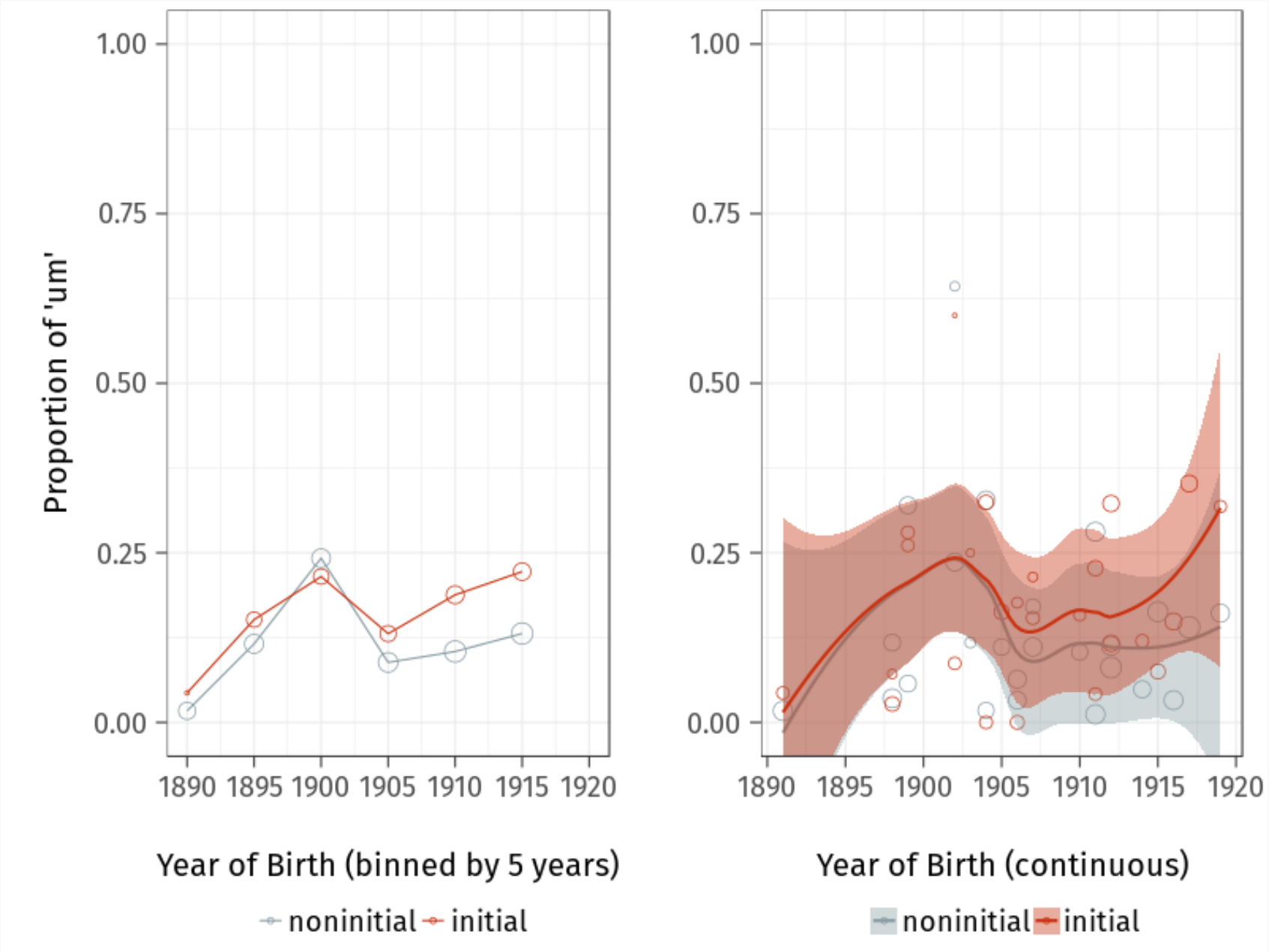
Proportional frequency in apparent time



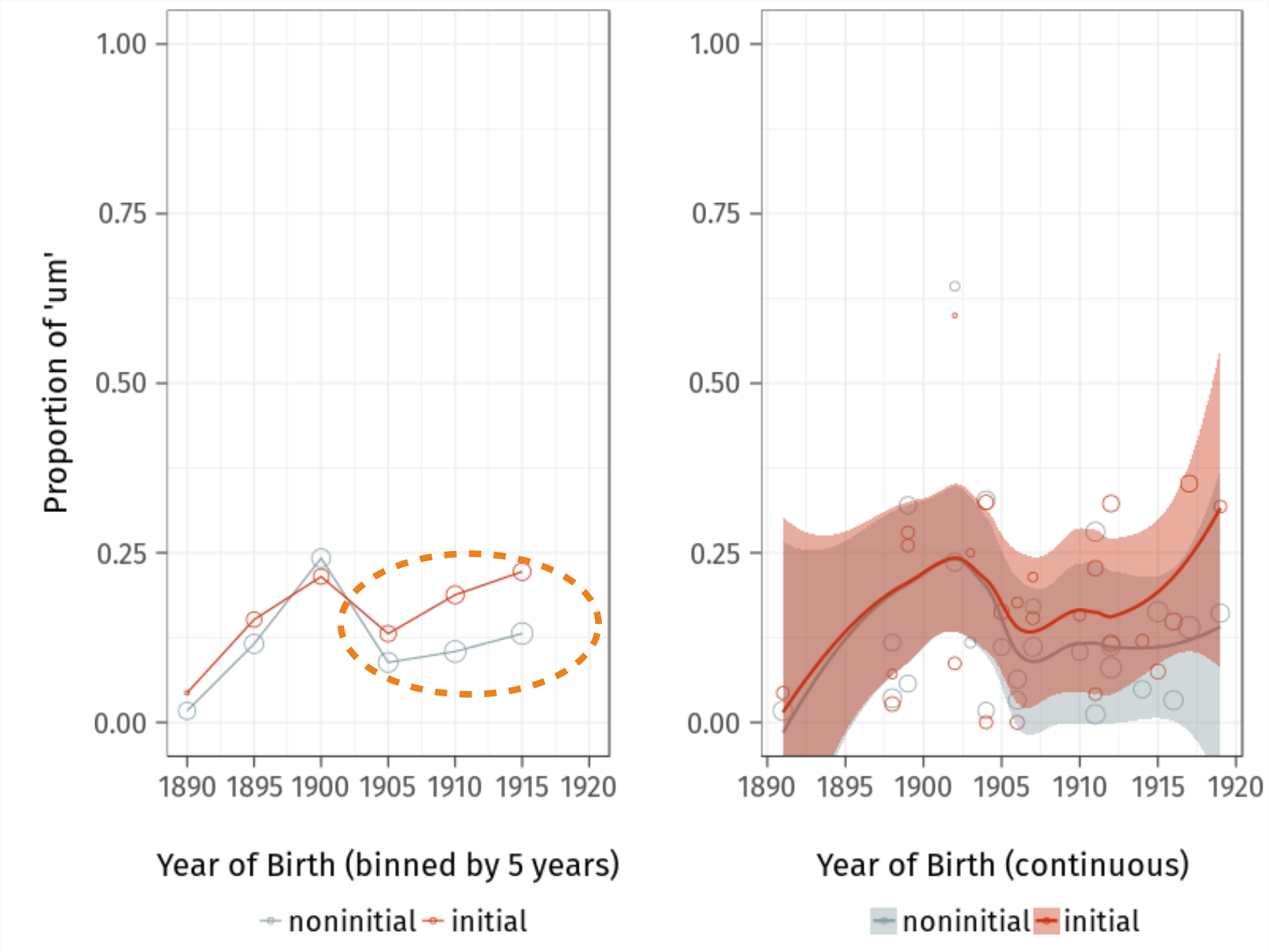
Proportional frequency in apparent time by gender



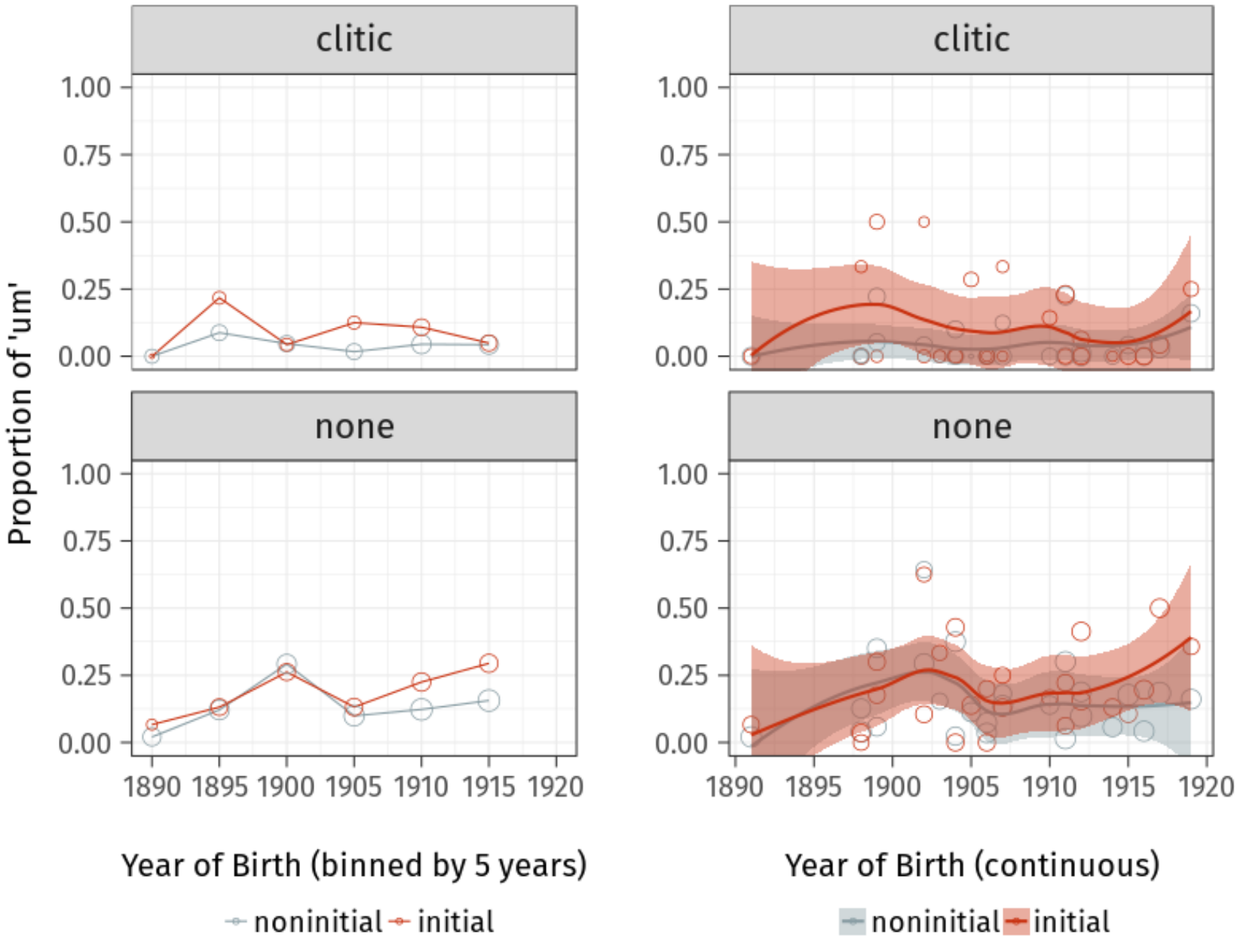
Proportional frequency in apparent time by position



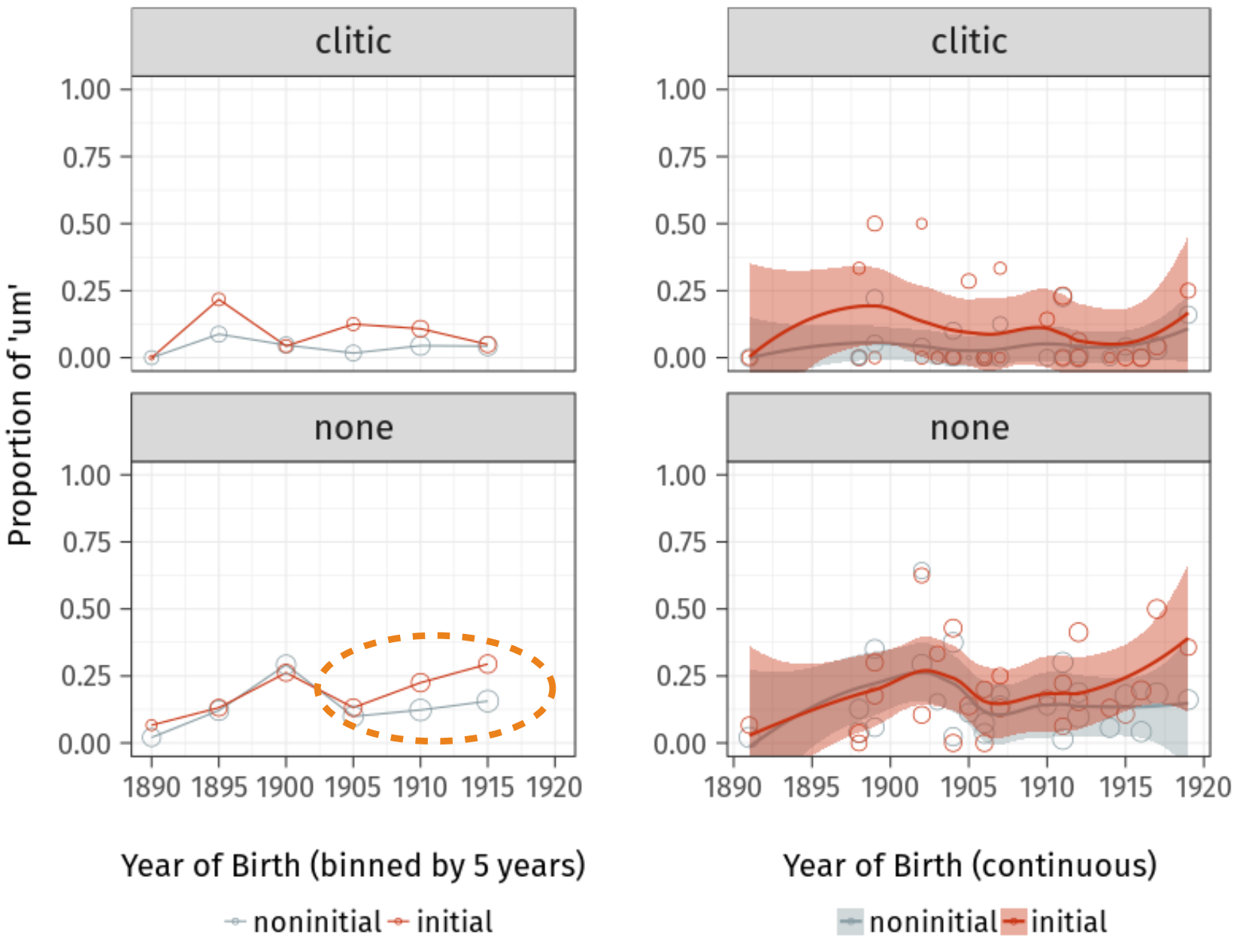
Proportional frequency in apparent time by position



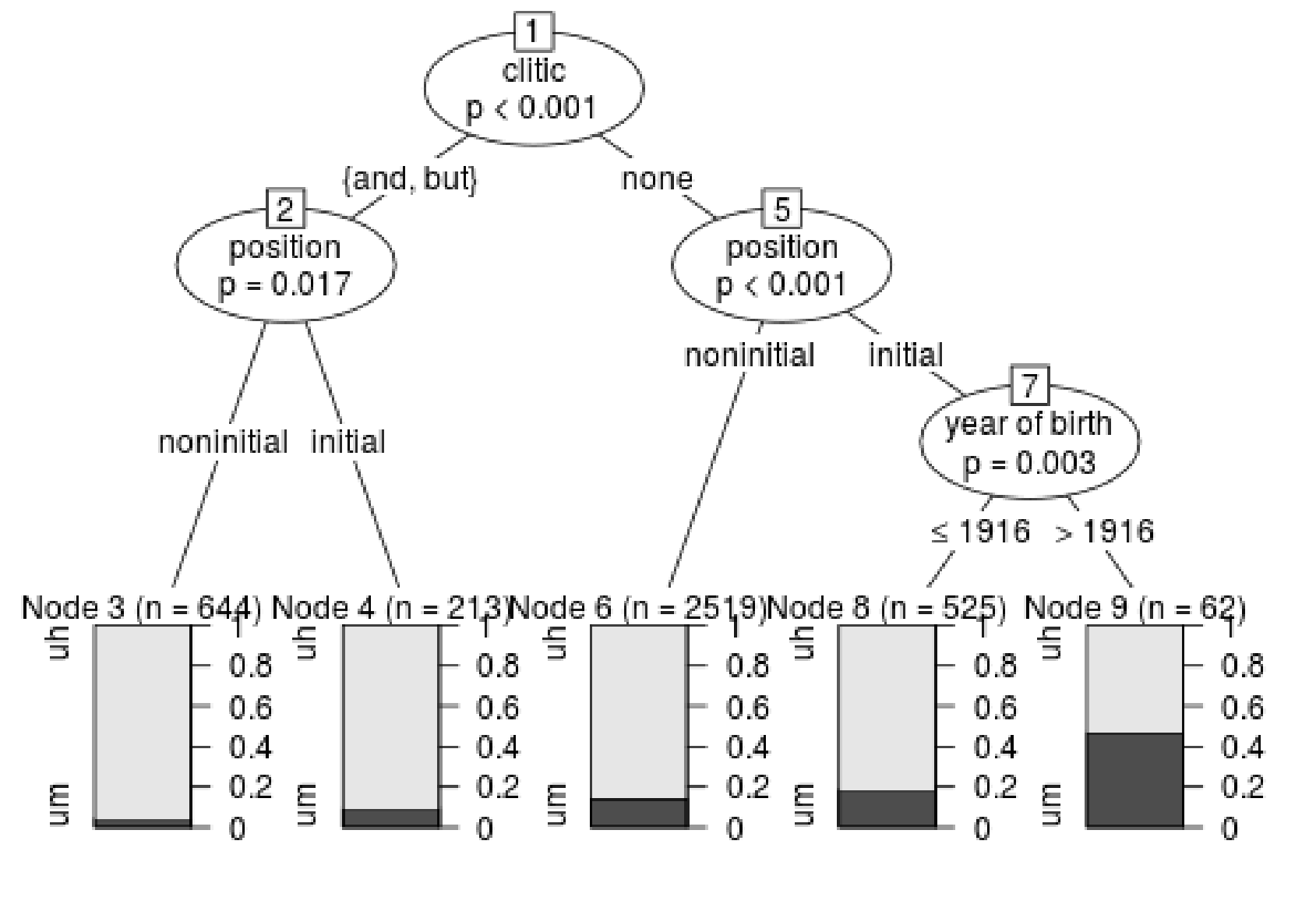
Proportional frequency in apparent time by position and clitic



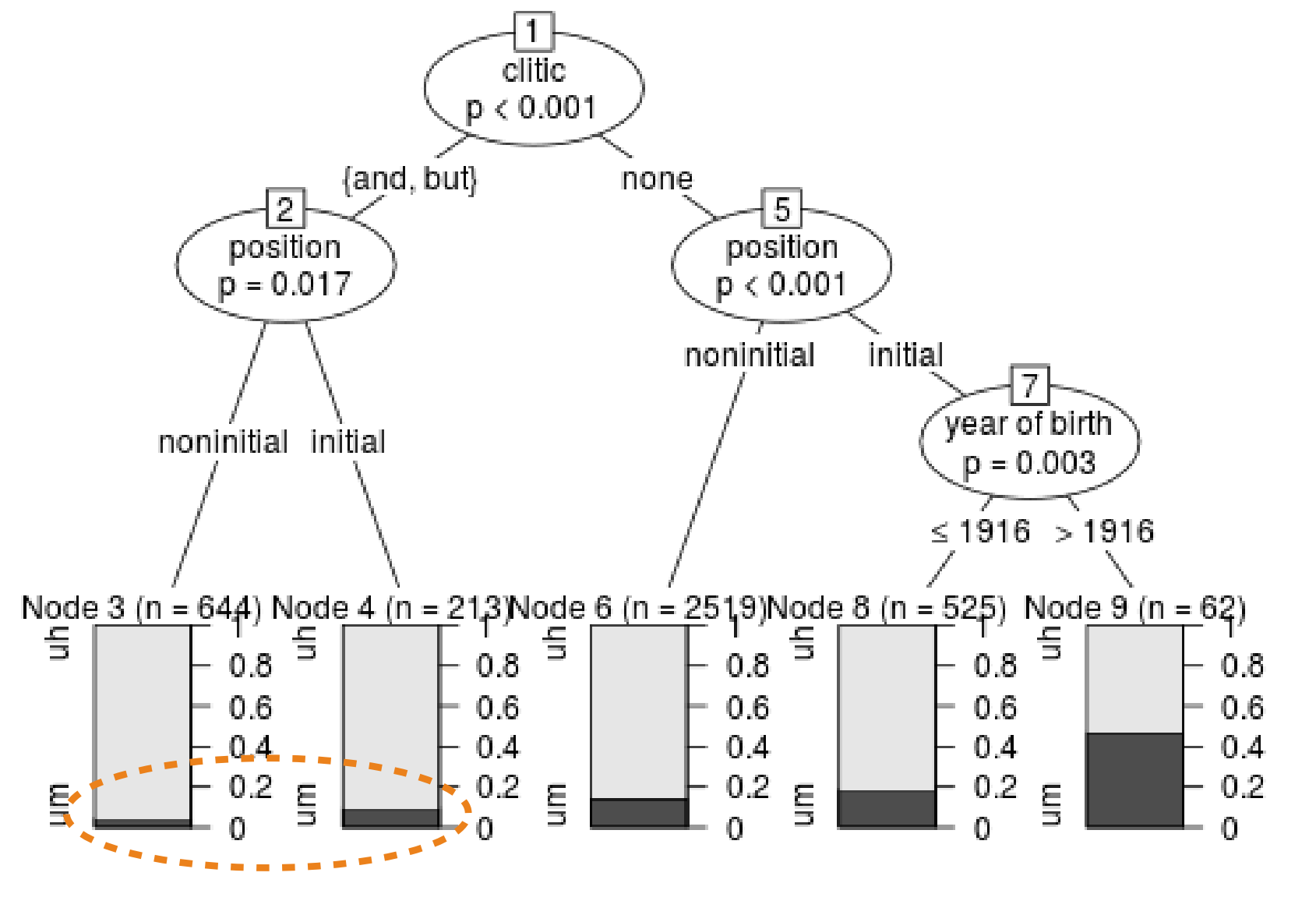
Proportional frequency in apparent time by position and clitic



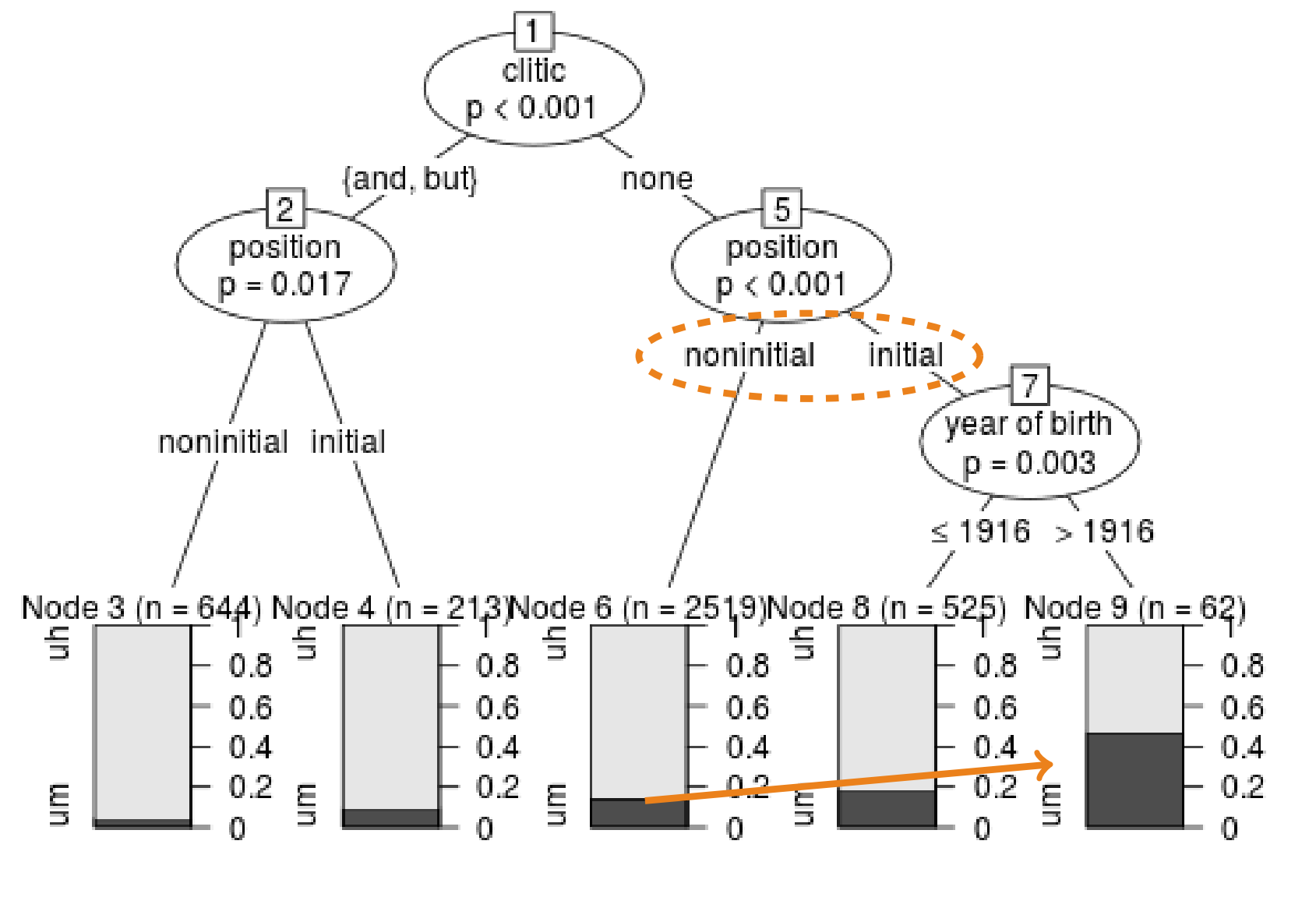
Conditional inference tree: farmers



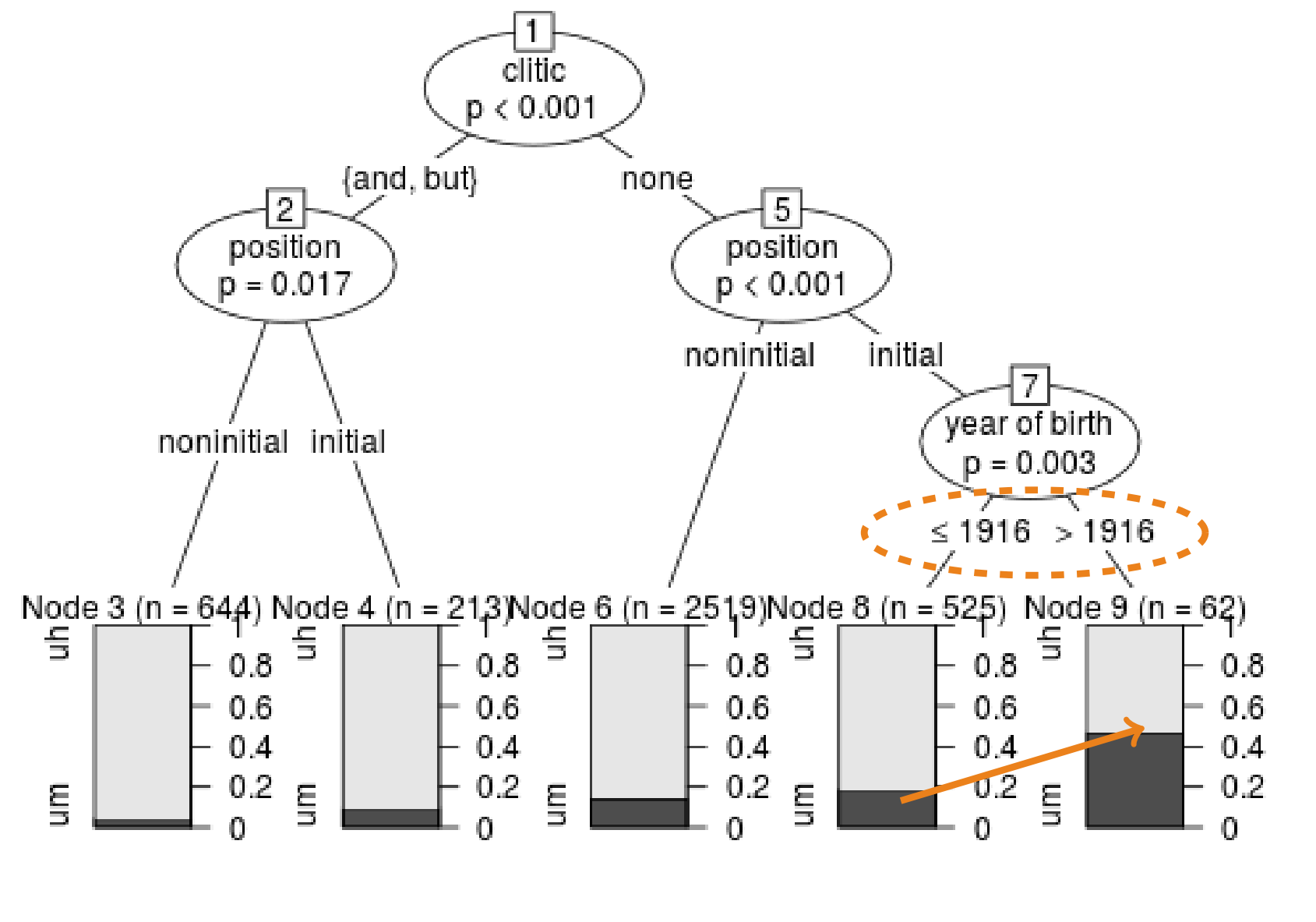
Conditional inference tree: farmers



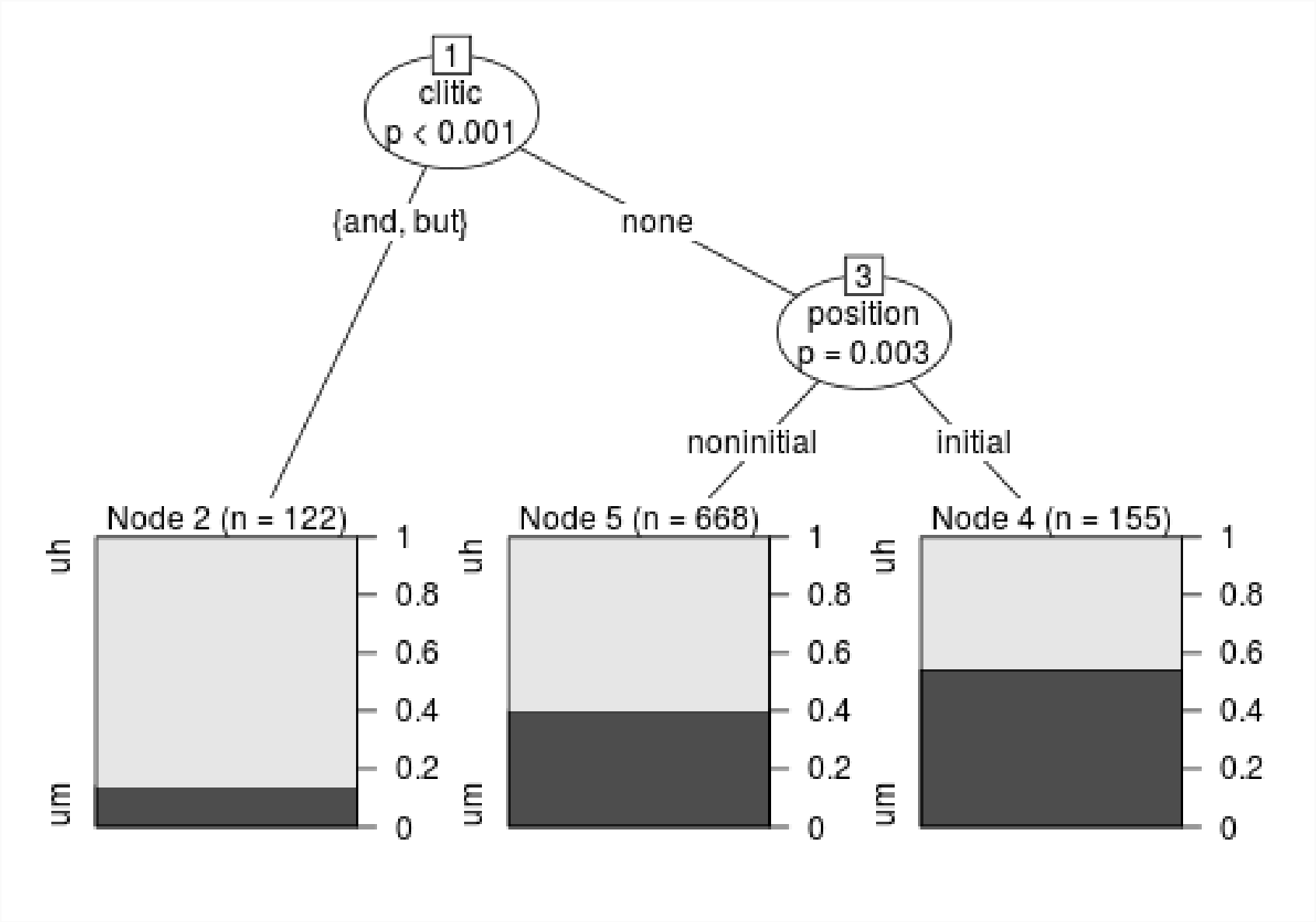
Conditional inference tree: farmers



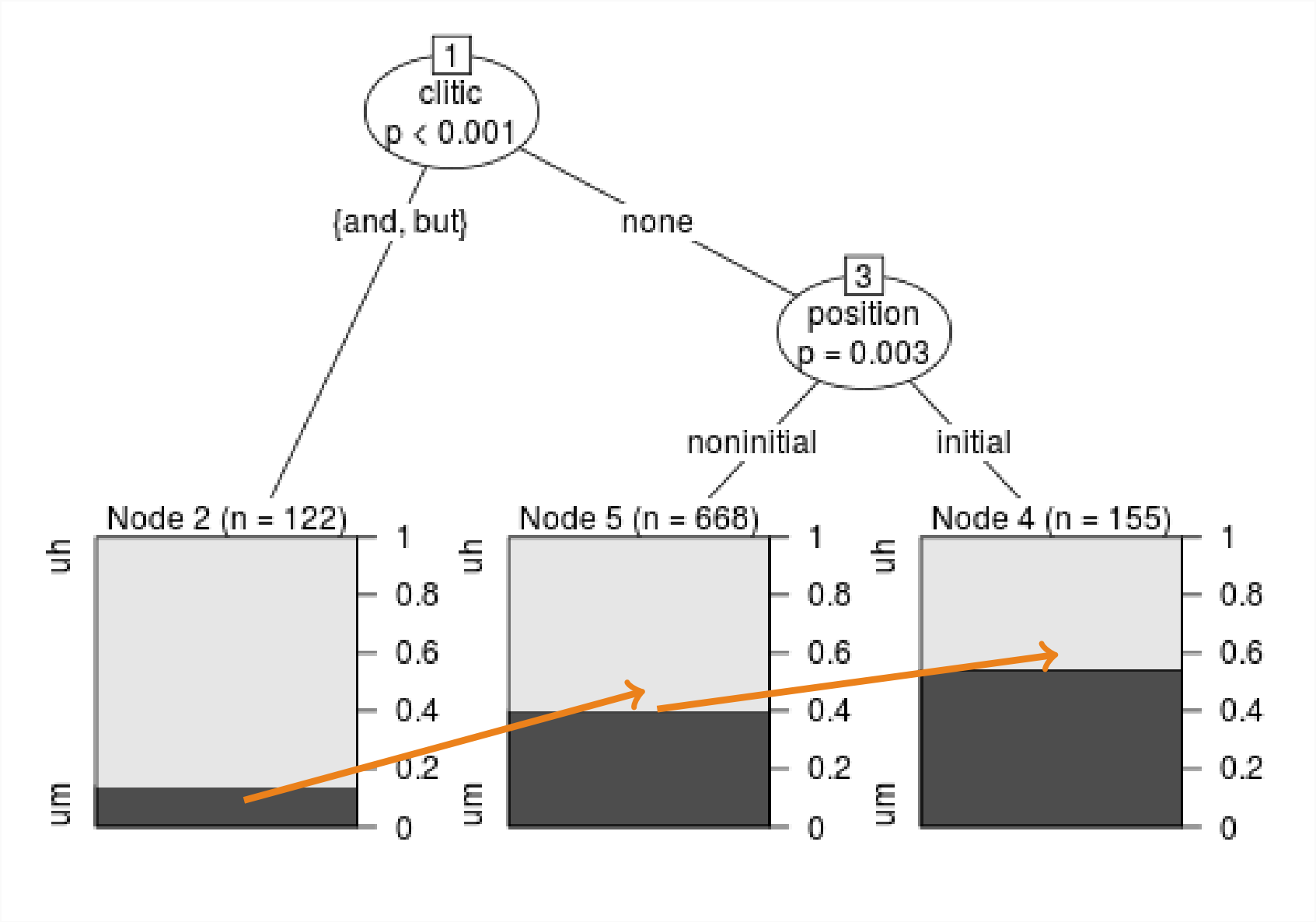
Conditional inference tree: farmers



Conditional inference tree: interviewers



Conditional inference tree: interviewers



Interim summary

- We see the beginning of the S-curve of change toward *um* in apparent time.
  - While other work has shown that women lead this change, in our data, older women actually use more *um* than the younger women.
  - (more on this later)
- Looking at the internal factors:
  - Cliticized forms favour *uh*
  - Some evidence of positional divergence (possibly consistent with a new utterance-initial discourse function that favours *um*, Fruehwald, 2016)
- Conditional inference trees confirm that internal constraints persist with the younger interviewers (though the baseline rate of *um* is higher).

Assessing functional expansion

- Fruehwald (2016) tests the hypothesis that **functional expansion** triggered the rise of *um* by considering changes to the relative frequency of variants over time.

Assessing functional expansion

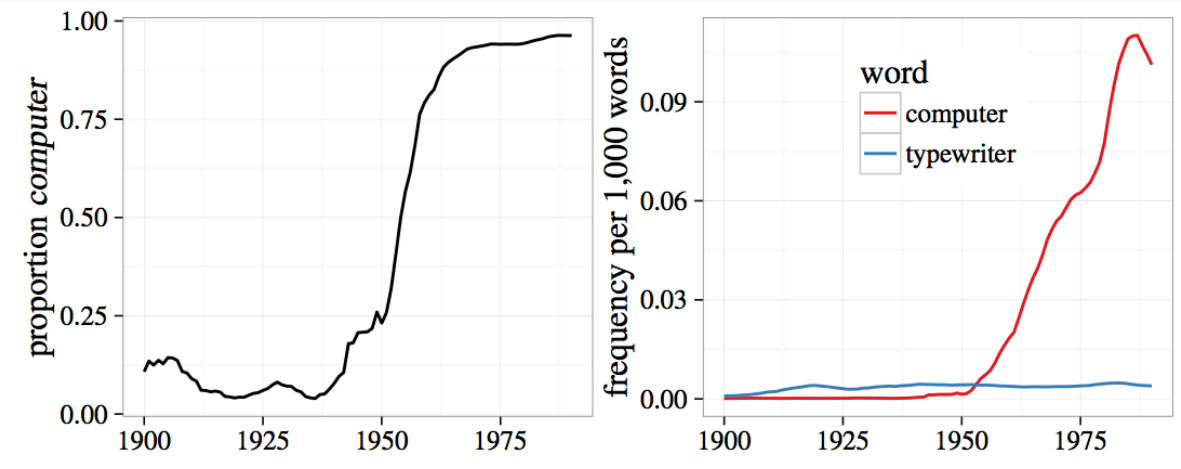
- Fruehwald (2016) tests the hypothesis that **functional expansion** triggered the rise of *um* by considering changes to the relative frequency of variants over time.
    - New discourse-pragmatic functions should be additive on the relative frequency.

Assessing functional expansion

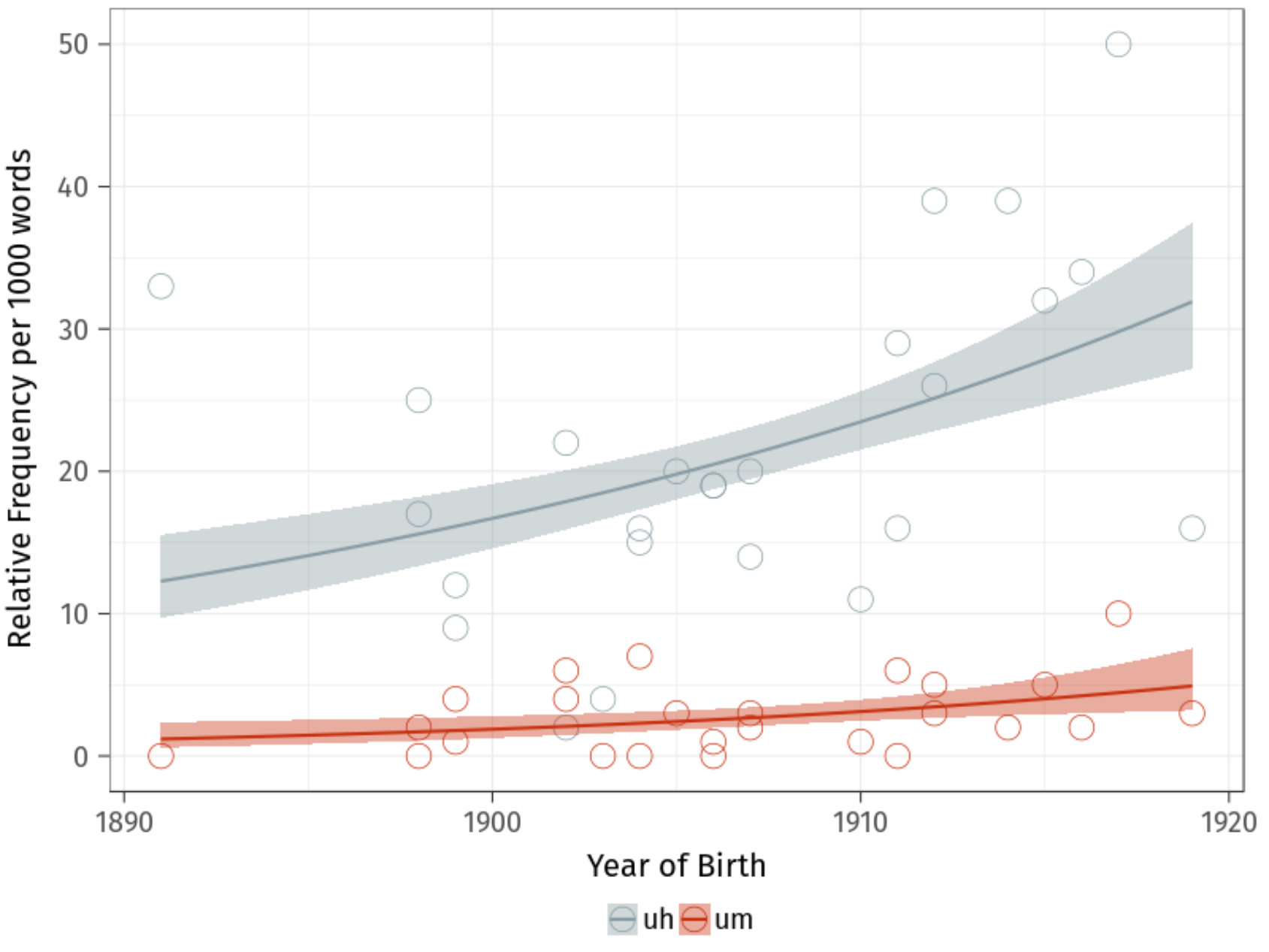
- Fruehwald (2016) tests the hypothesis that **functional expansion** triggered the rise of *um* by considering changes to the relative frequency of variants over time.
  - New discourse-pragmatic functions should be additive on the relative frequency.
  - If the new function is restricted to one variant, the relative frequency of one variant is expected to rise.

Assessing functional expansion

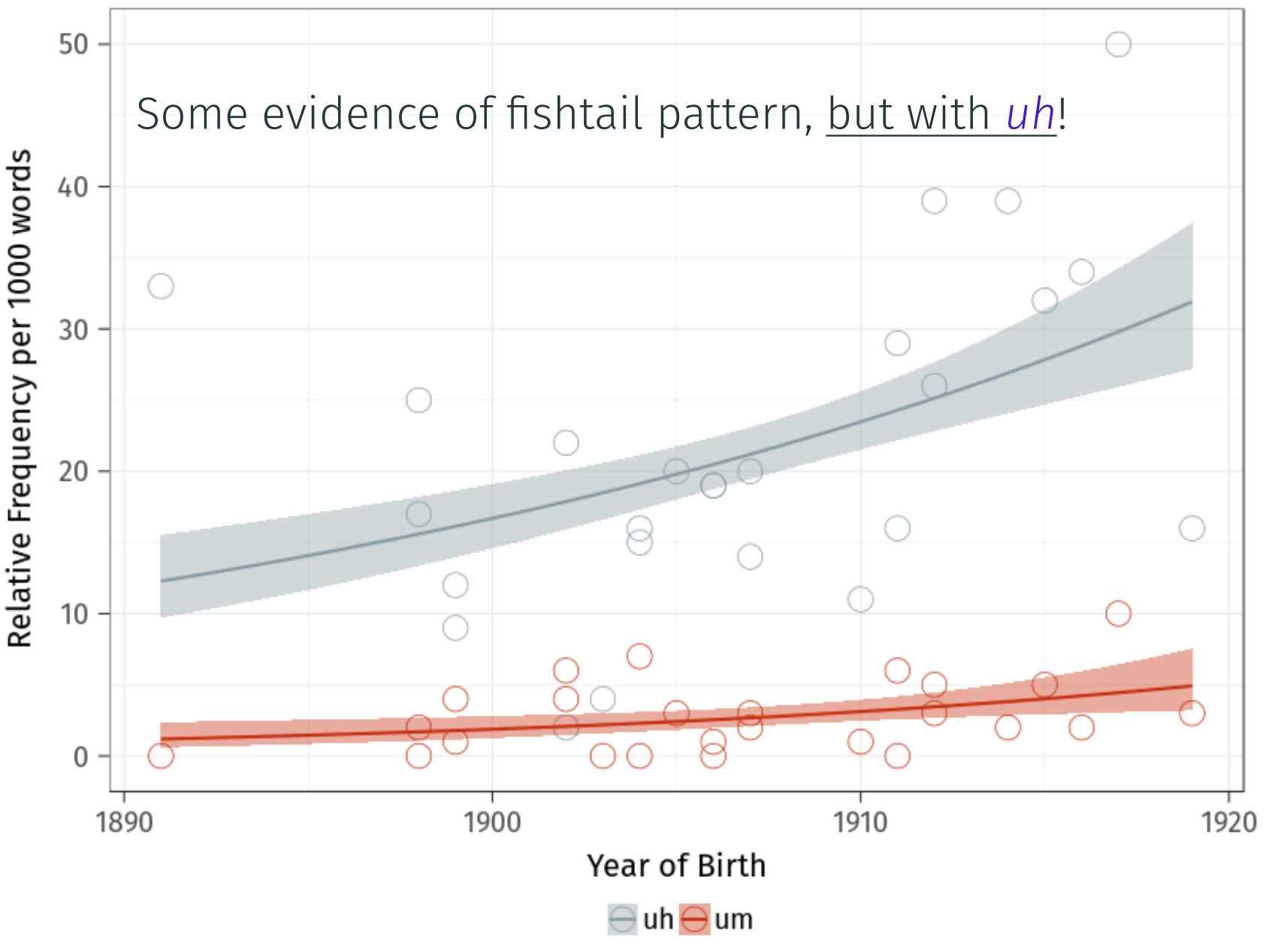
- Fruehwald (2016) tests the hypothesis that **functional expansion** triggered the rise of *um* by considering changes to the relative frequency of variants over time.
  - New discourse-pragmatic functions should be additive on the relative frequency.
  - If the new function is restricted to one variant, the relative frequency of one variant is expected to rise.
  - Expect a **fishtail pattern** as with *computer* and *typewriter* over time: once *computer* gained its contemporary meaning, its relative frequency grew additively as that meaning became more frequent.



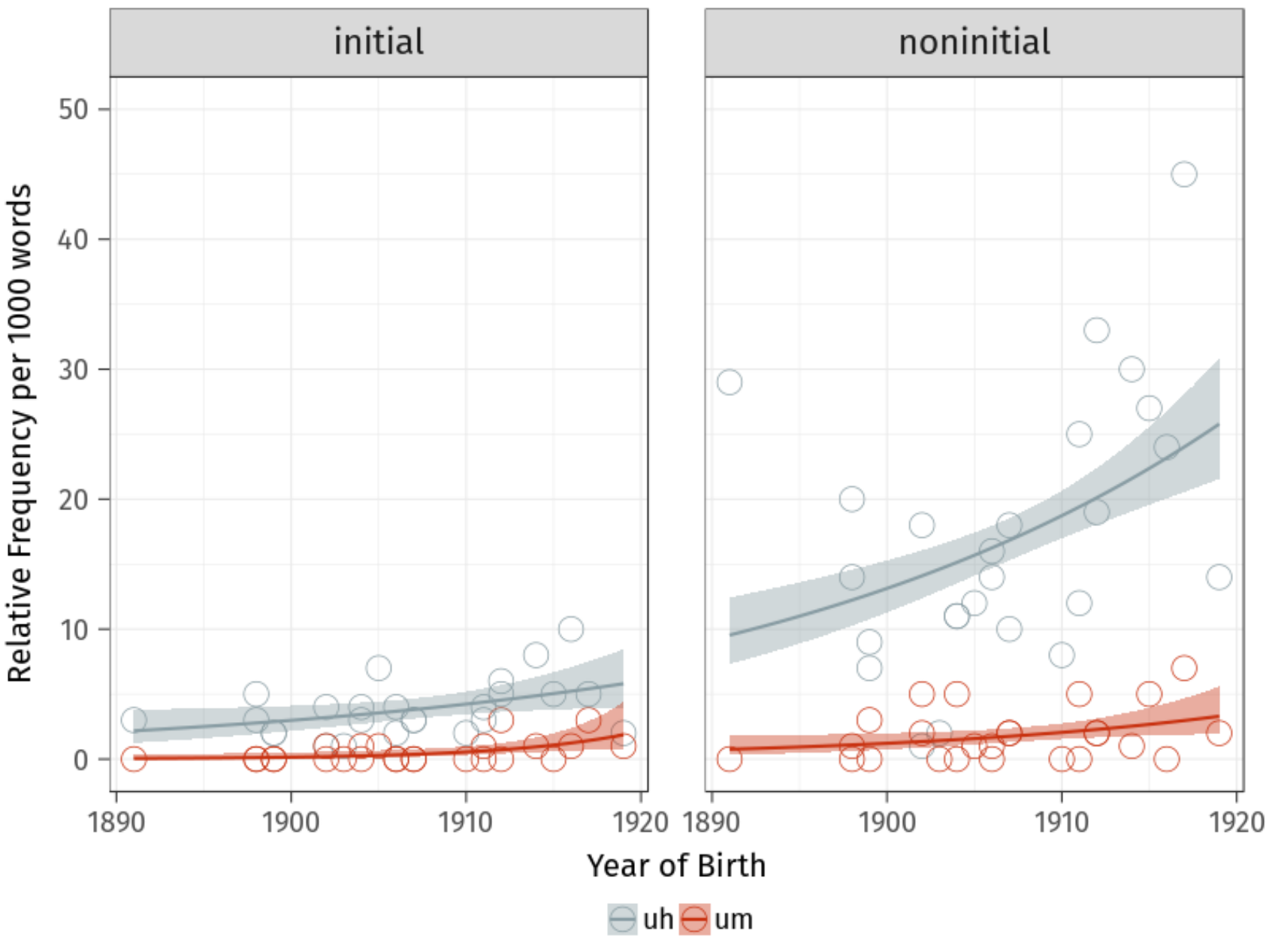
Relative frequency



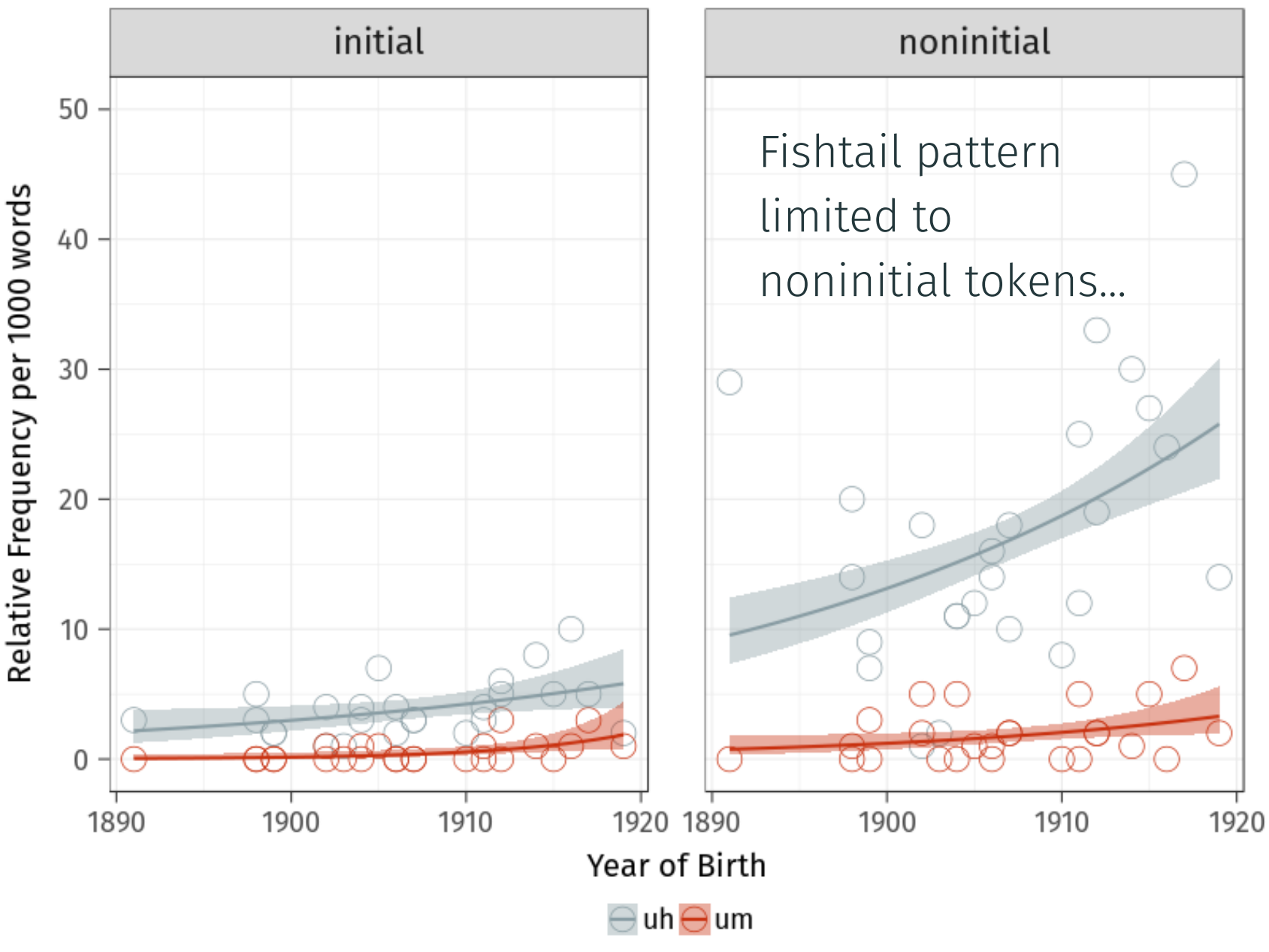
Relative frequency



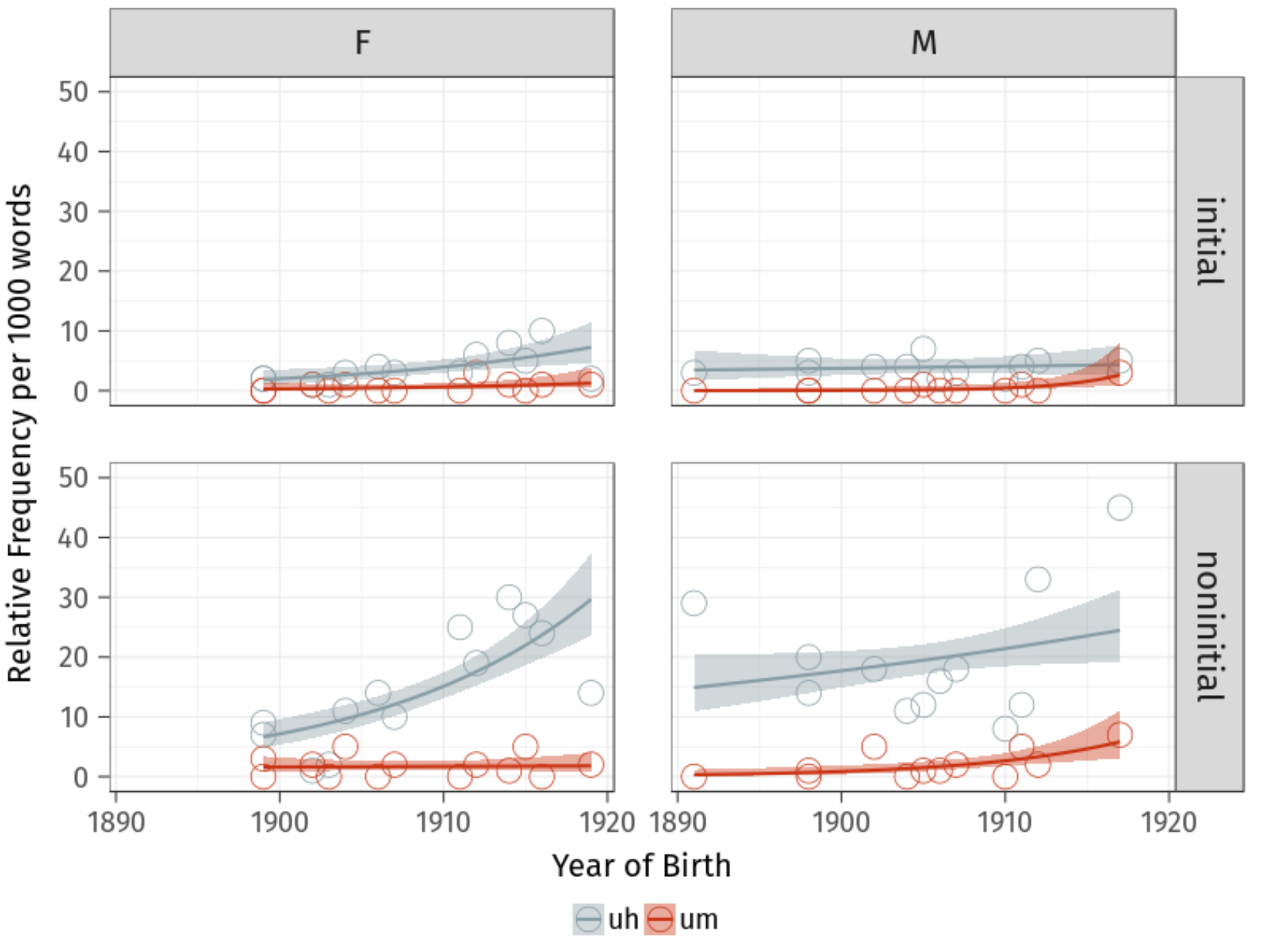
Relative frequency by position



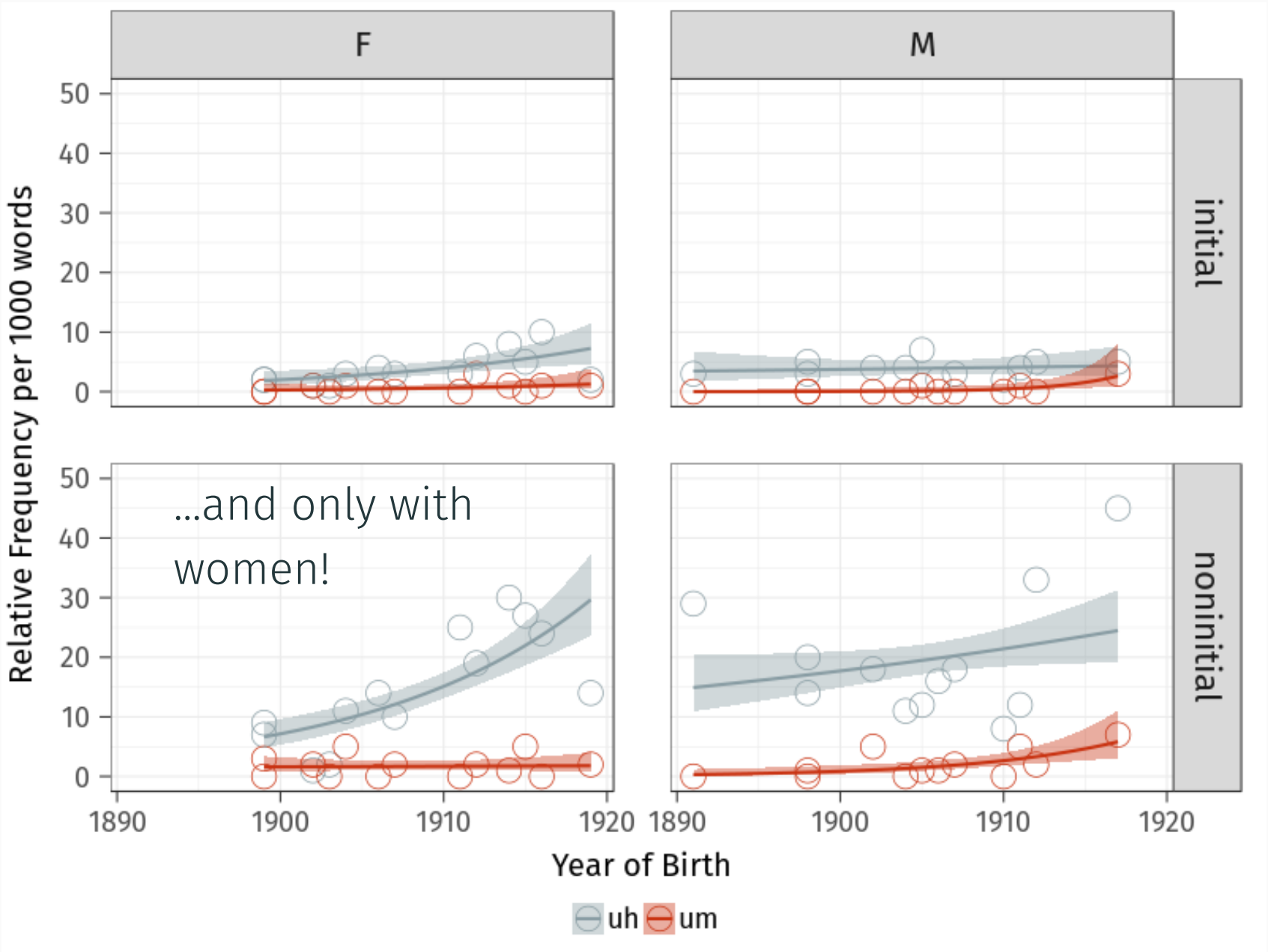
Relative frequency by position



Relative frequency by gender and position



Relative frequency by gender and position



Contrasting two speakers

High UHM user (b. 1917)

INT: And what types of fruit . did you grow?

NO-11: Well the uh . originally uh when they came- uh grandfather bought the property in nineteen hundred and uh . um . to begin with there was very- there were very few fruit trees on it and they planted .uh . our orchard of uh . peaches. And uh waiting- while they waited for the peaches to come into bearing, they planted raspberries between the rows, so it started out as principally a raspberry farm I suppose but . it evolved into a farm that uh principally grew peaches and cherries mainly sweet cherries.

Contrasting two speakers

High UHM user (b. 1917)

INT: And what types of fruit . did you grow?

NO-11: Well the uh . originally uh when they came- uh grandfather bought the property in nineteen hundred and uh . um . to begin with there was very- there were very few fruit trees on it and they planted .uh . our orchard of uh . peaches. And uh waiting- while they waited for the peaches to come into bearing, they planted raspberries between the rows, so it started out as principally a raspberry farm I suppose but . it evolved into a farm that uh principally grew peaches and cherries mainly sweet cherries.

Low UHM user (b. 1903)

INT: Okay. And how much . older was the very oldest?

NO-36: The oldest was born … in eighteen ninety two … and then my sister Lianne, eighteen ninety four … Greg, eighteen ninety eight . Sally nineteen hundred and one … I was born nineteen hundred and three . and that's it.

INT: Okay, and how old was your dad when you were born? At-

NO-36: … I- … how old was my dad when I was born? Oh.

INT: I think we had figured out that he was probably somewhere around forty five.

NO-36: Oh yes.

INT: And your mom was?

NO-36: Thirty . five?

INT: Thirty- oh-

NO-36: Is that it?

INT: Yup. Good.

Conclusion

Conclusion

- Importance of looking at discourse-pragmatic variation from multiple quantitative angles.
  - Conflicting views on functional expansion.
- Exploring data from before the rise of *um* hasn't offered us an explanation for its dramatic rise.
  - Fruehwald (2016) didn't find an effect of position either.
  - The actuation problem remains elusive!
- But when we do find suggestive evidence of a different change—seems to be a change from unfilled pauses to filled pauses, *most of which are being filled by* uh
- Something for future investigation!

References I

Clark, H. H. & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73–111.

Denis, D. (2016). Oral histories as a window to sociolinguistic history and language history: Exploring earlier Ontario English with the Farm Work and Farm Life Since 1890 oral history collection. *American speech*, *91*(4), 513–516.

Fruehwald, J. (2016). Filled pause choice as a sociolinguistic variable. *University of Pennsylvania Working Papers in Linguistics*, *22*(2), 6.

Gadanidis, T. (2018). Um, about that, uh, variable: *Uh* and *um* in teen instant messaging. Paper presented at DiPVaC4, Helsinki, Finland.

References II

Levelt, W. J. (1989). *Speaking: from intention to articulation*. MIT Press Series in Natural-Language Processing. MIT Press, Cambridge, Massachusetts.

Maclay, H. & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, *15*(1), 19–44.

Pichler, H. (2010). Methods in discourse variation analysis: reflections on the way forward. *Journal of Sociolinguistics*, *14*(5), 581–608.

Tottie, G. (2016). Planning what to say: uh and um among the pragmatic markers. In G. Kaltenböck, E. Keizer, & A. Lohmann (Eds.), *Outside the Clause: Form and function of extra-clausal constituents* (pp. 97–122). John Benjamins Publishing Company.

References III

Tottie, G. (2017). From pause to word: *uh*, *um* and *er* in written American English. *English Language & Linguistics*, 1–26.

Wieling, M., Grieve, J., Bouma, G., Fruehwald, J., Coleman, J., & Liberman, M. (2016). Variation and change in the use of hesitation markers in Germanic languages. *Language Dynamics and Change*, *6*(2), 199–234.

Acknowledgements

- Archives of Ontario
- University of Toronto Open Grant, 2017 – 2019
- SGS Conference Grant & Linguistics Department SIG travel grant