

# No country for old members: User lifecycle and linguistic change in online communities

Cristian Danescu-Niculescu-Mizil

Robert West

Dan Jurafsky

Jure Leskovec

Christopher Potts

Stanford University

Max Planck Institute SWS

# User-community relation

## Our high-level goal:

Analyzing the relation between an user and an online community



# User-community relation

## Our high-level goal:

Analyzing the relation between an user and an online community



Joining

# User-community relation

**Our high-level goal:**

Analyzing the relation between an user and an online community



Joining →

# User-community relation

## Our high-level goal:

Analyzing the relation between an user and an online community



Joining →

Abandoning

# User-community relation

**Concrete questions we address:**

→ How does a user become member of a community?

# User-community relation

## Concrete questions we address:

- How does a user become member of a community?
- How do user & community practices co-evolve?

# User-community relation

## Concrete questions we address:

- How does a user become member of a community?
- How do user & community practices co-evolve?
- Can we predict when a user will leave the community?



# Main insight: linguistic change

Language practices (norms, etiquette, ...)

→ build collective identity

→ foster individual expression

# Main insight: linguistic change

Language practices (norms, etiquette, ...)

→ build collective identity

→ foster individual expression

**Linguistic change** allows us to capture

→ **relation between members and their community**

# Our approach: **linguistic change**

→ **Statistical framework for tracking linguistic change**

# Our approach: **linguistic change**

- **Statistical framework for tracking linguistic change**
- **Measures of user reaction to linguistic change**

# Our approach: **linguistic change**

- **Statistical framework for tracking linguistic change**
- **Measures of user reaction to linguistic change**
- **Features predicting when user will leave the community**

# Longitudinal data

Complete linguistic record of **two** online communities:



Community starts

Present

complete linguistic record

data available at <http://snap.stanford.edu/data/>

# Longitudinal data

Complete linguistic record of **two** online communities:

## Beeradvocate



2001 (inception)

2011

10 years of complete linguistic record

1,600,000 posts

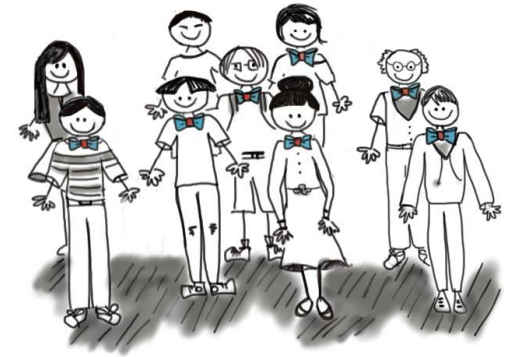
33,000 users

data available at <http://snap.stanford.edu/data/>

# Longitudinal data

Complete linguistic record of **two** online communities:

ratebeer



2001 (inception)

2011

10 years of complete linguistic record

3,000,000 posts

30,000 users

data available at <http://snap.stanford.edu/data/>



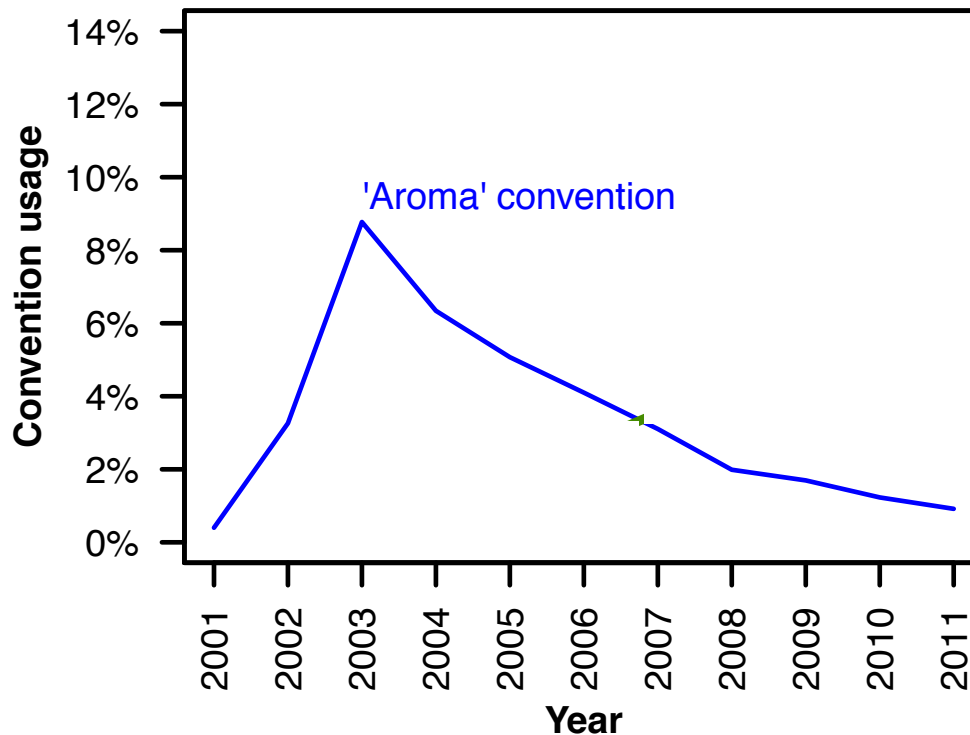
# Linguistic change: one example

# Linguistic change: one example

... **Aroma:** Buttery, slightly spicy malt notes ...

# Linguistic change: one example

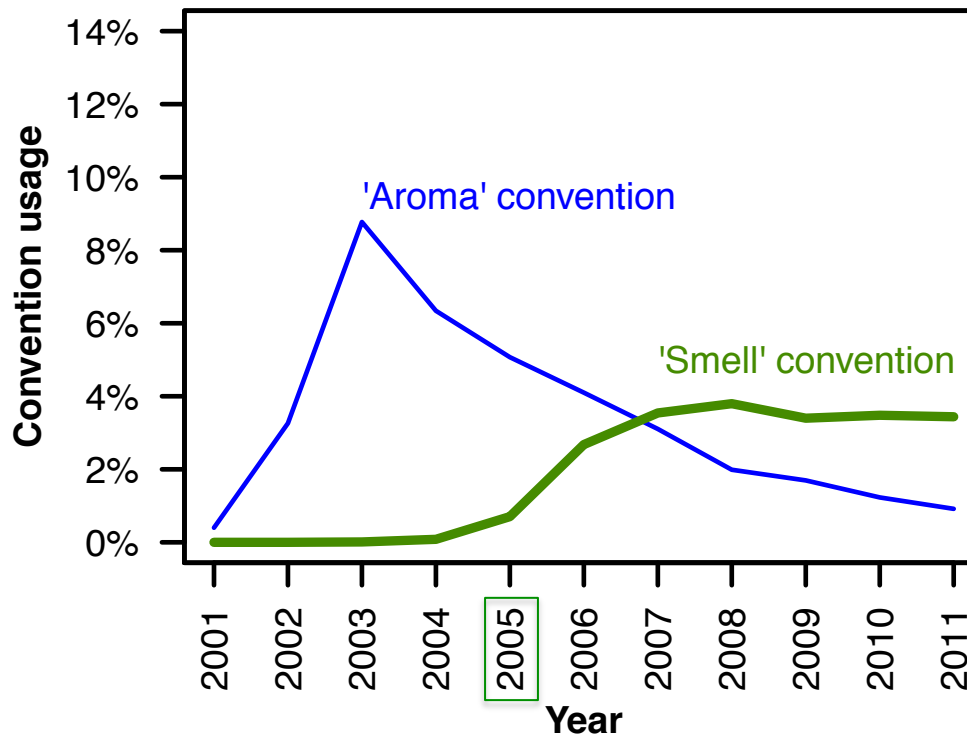
... **Aroma**: Buttery, slightly spicy malt notes ...



# Linguistic change: one example

... **Aroma**: Buttery, slightly spicy malt notes ...

... **S**: Great nose of ginger, honey, perfume ...

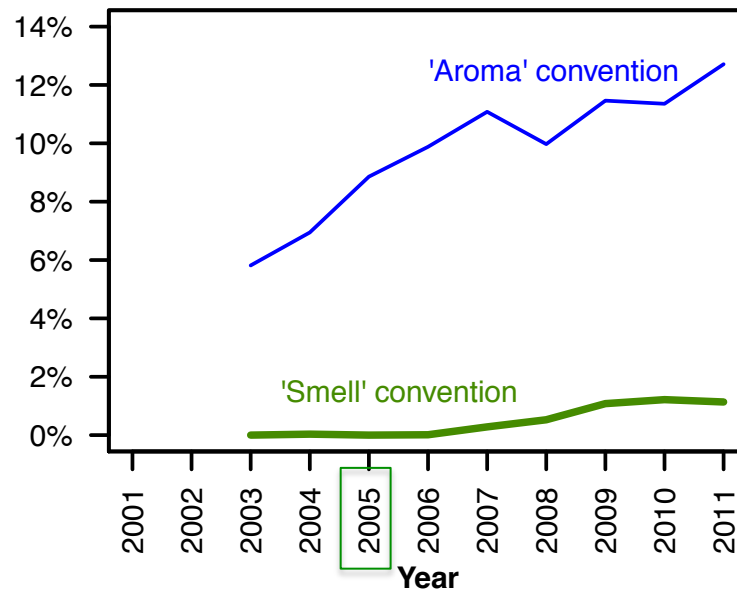


# Linguistic change: a puzzle

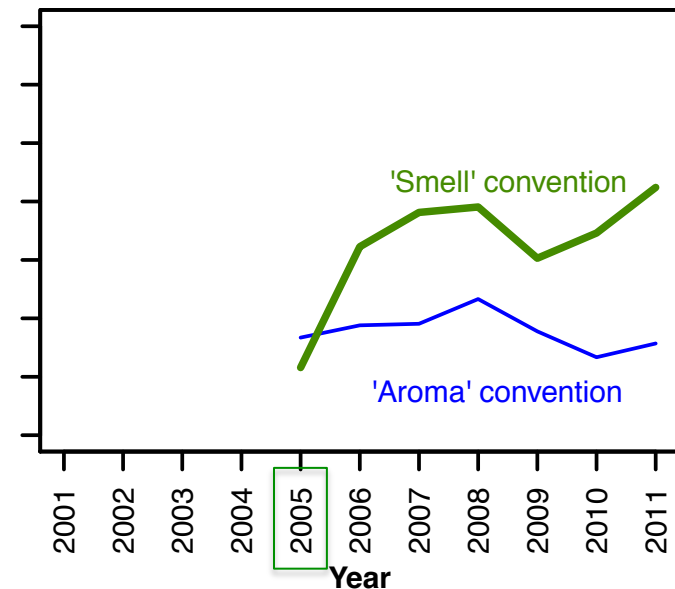
... **Aroma**: Buttery, slightly spicy malt notes ...

... **S**: Great nose of ginger, honey, perfume ...

Users joining in 2003



Users joining in 2005



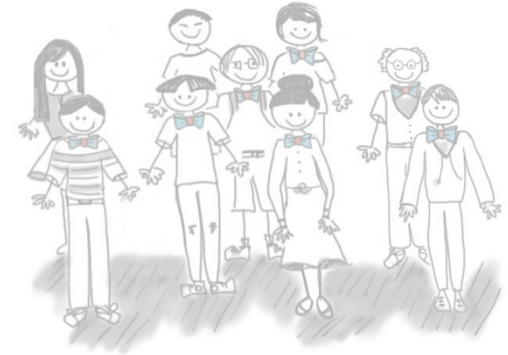
# Community-level change and user-level change



Other examples of community-level changes:

Re-tweet convention on Twitter, slang in hip-hop forums  
[Romero et al. 2011; Kooti et al. 2012; Garley and Hockenmaier 2012; inter alia]

# Community-level change and user-level change



2001

2011

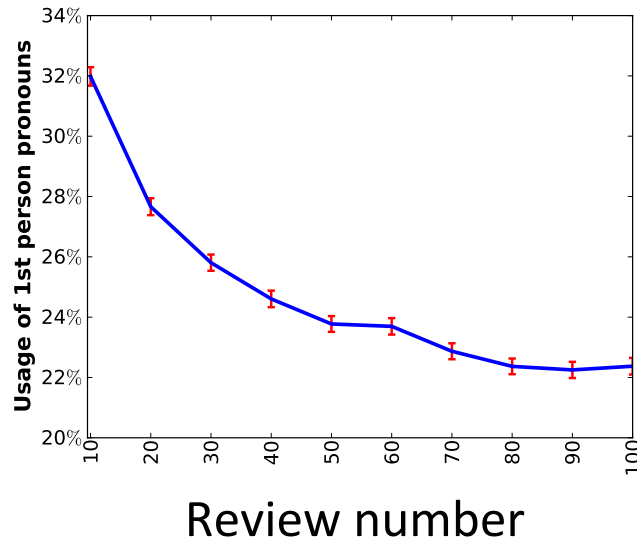


Joining

Abandoning

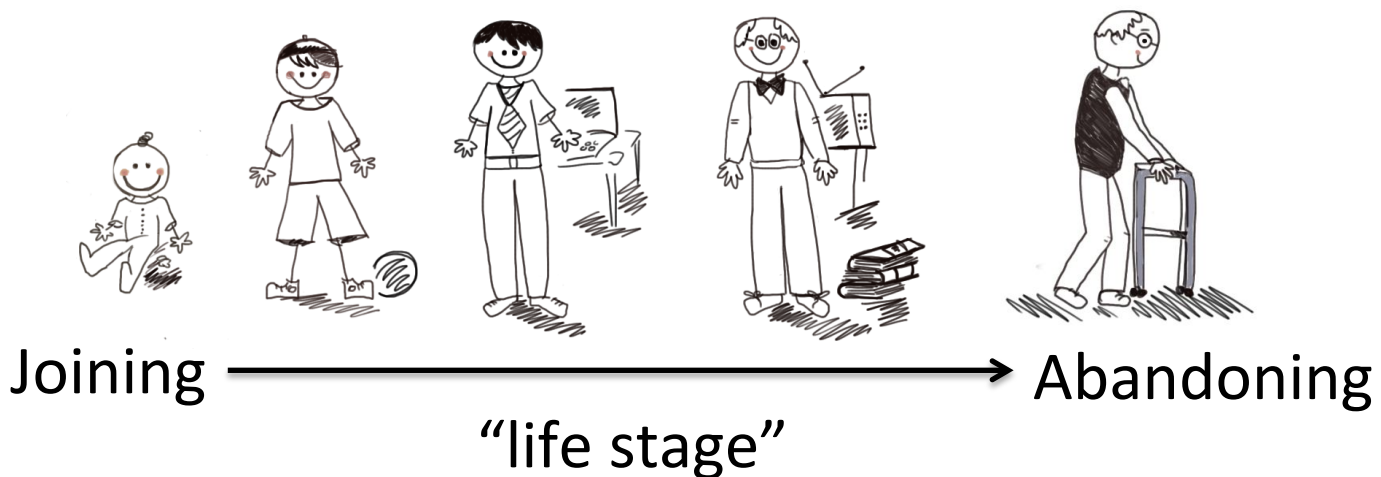
“life stage”

# Community-level change and user-level change



Example of user-level change:  
Decrease in usage of 1<sup>st</sup> person pronouns  
(e.g., I, me, mine, myself)

A sign of increasing identification with the  
community [Pennebaker 2007; Sherblom 2009]





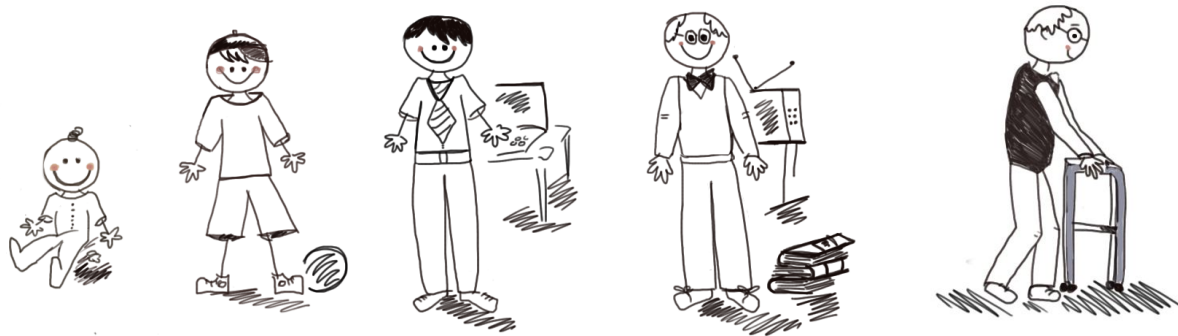
# Community-level change and user-level change

The rest of this talk: relation between these two levels of change



2001

2011



Joining

Abandoning

“life stage”

# Distance from the community

# Distance from the community



2001

2011

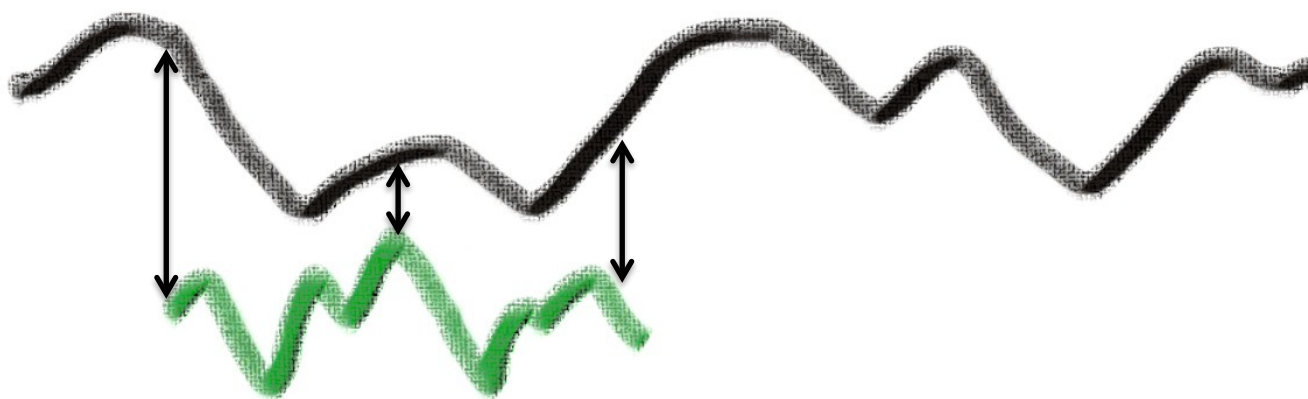


# Distance from the community



2001

2011



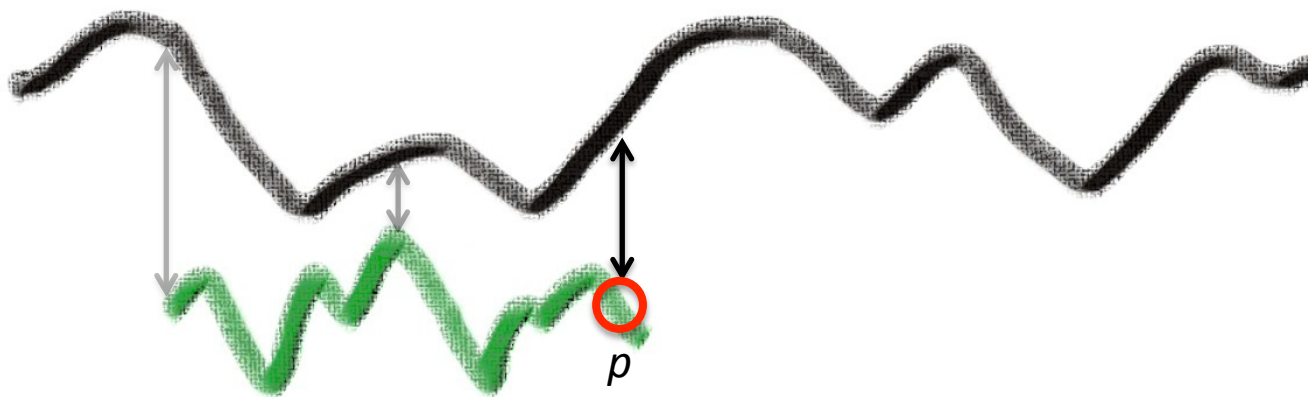
# Distance from the community



JAN 2006

2001

2011



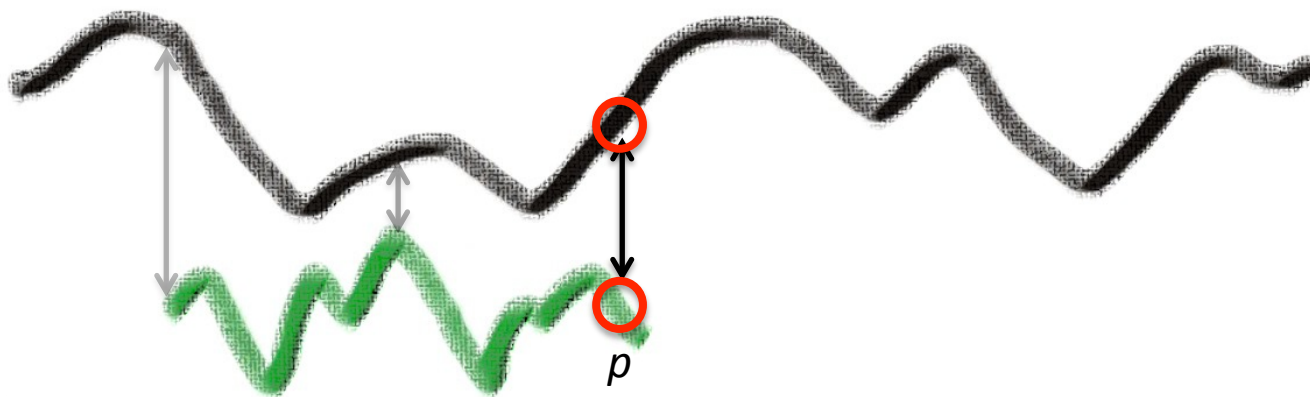
# Distance from the community



JAN 2006

2001

2011



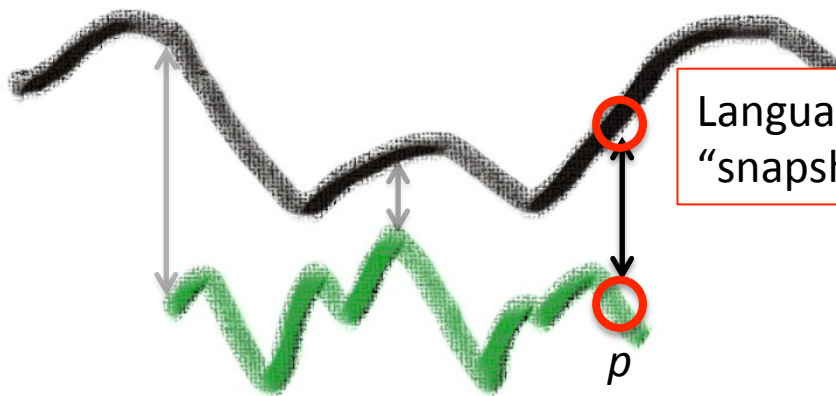
# Distance from the community



JAN 2006

2001

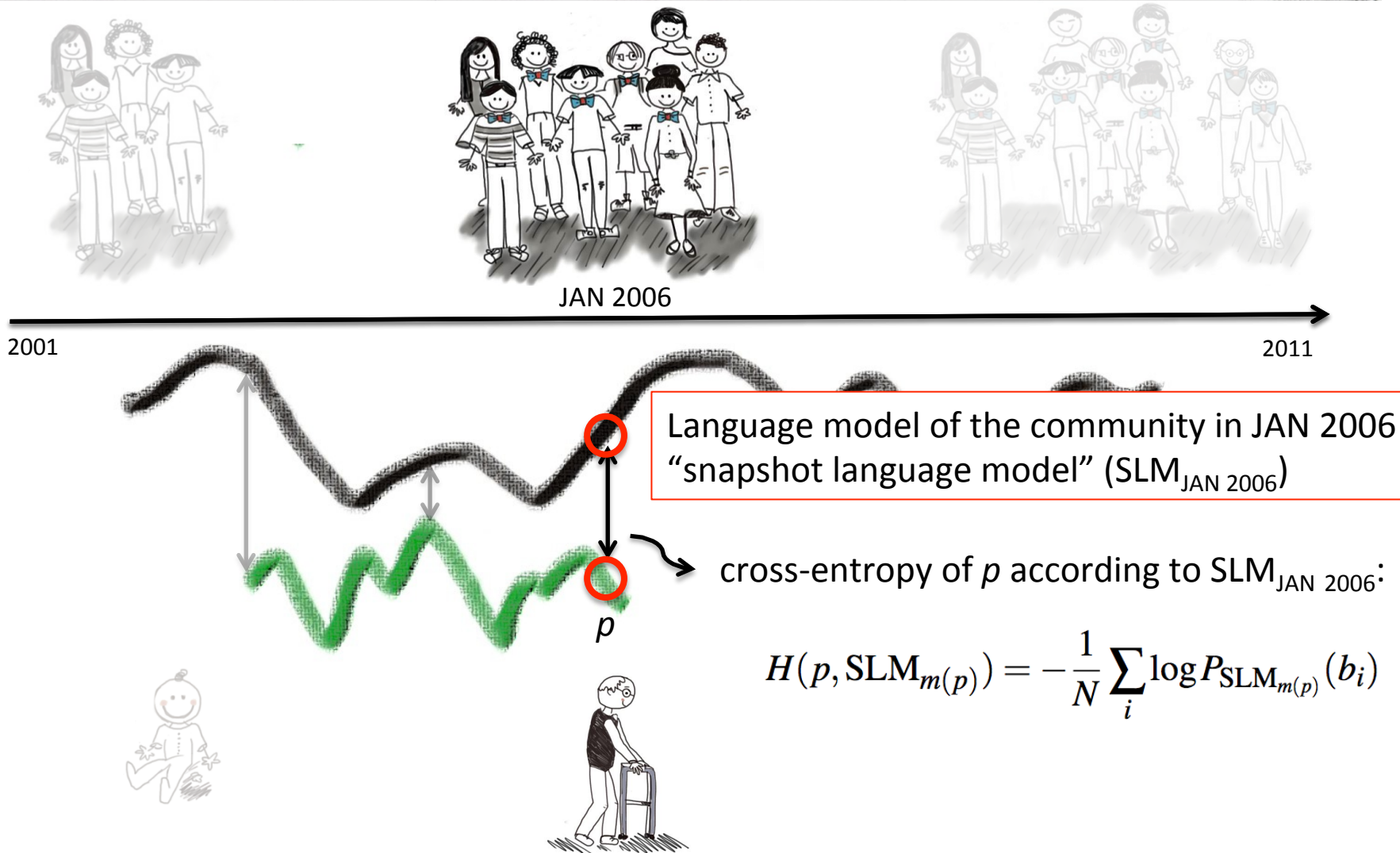
2011



Language model of the community in JAN 2006  
“snapshot language model” ( $SLM_{JAN\ 2006}$ )

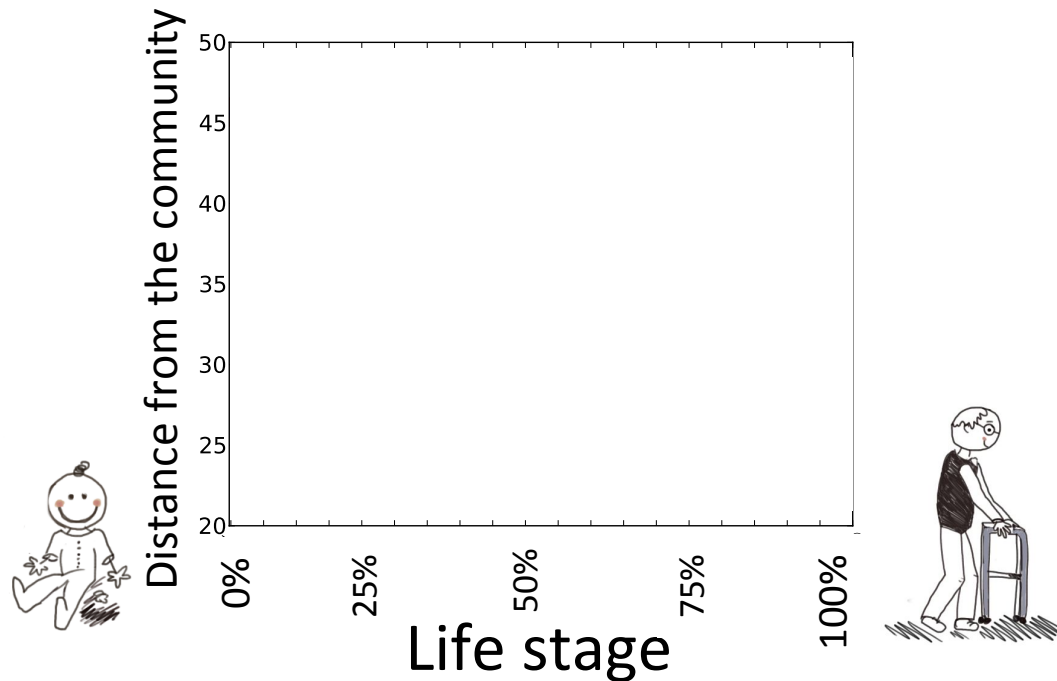


# Distance from the community



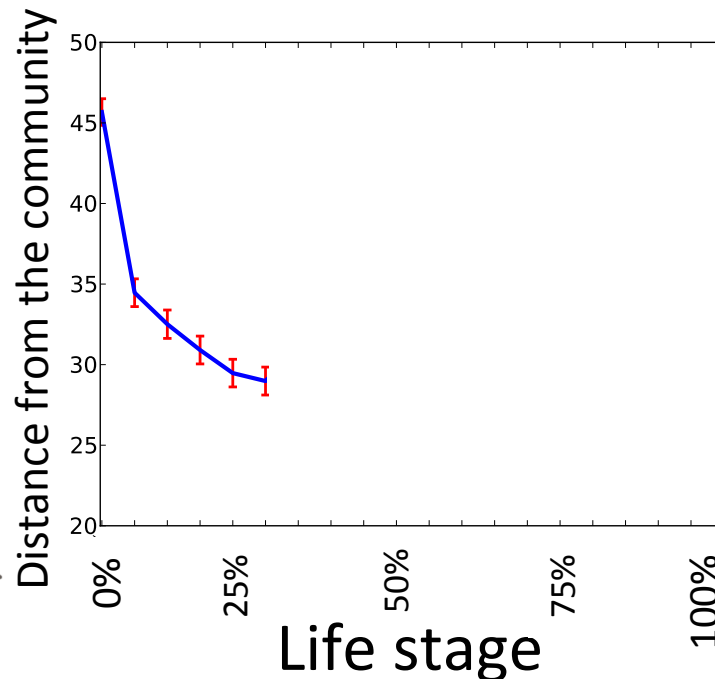


# Distance from the community

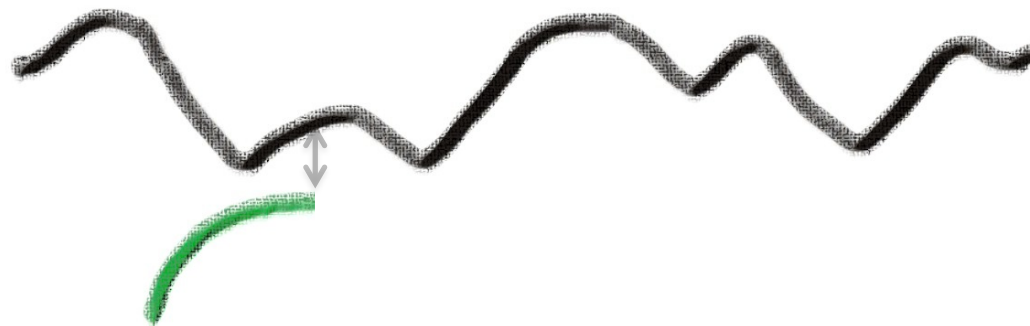


# Distance from the community

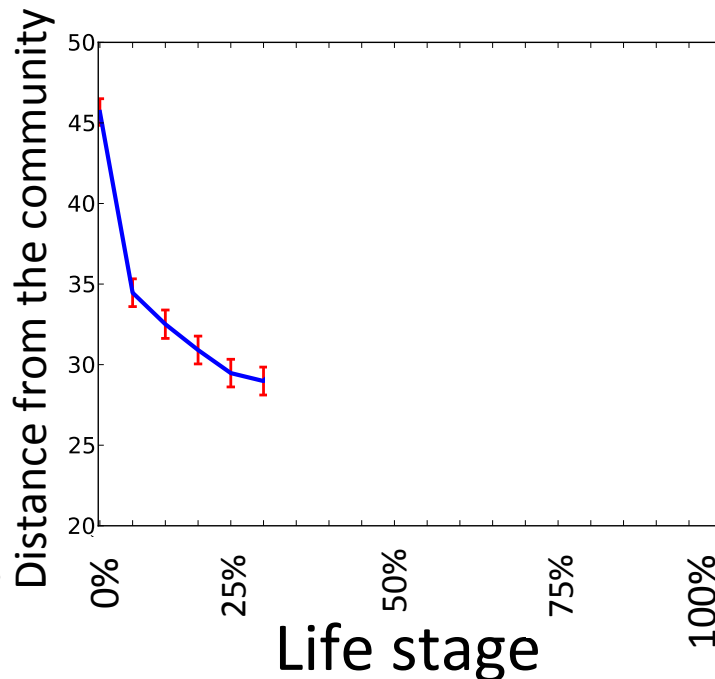
Stage 1:  
user **assimilates**  
the language of  
the community



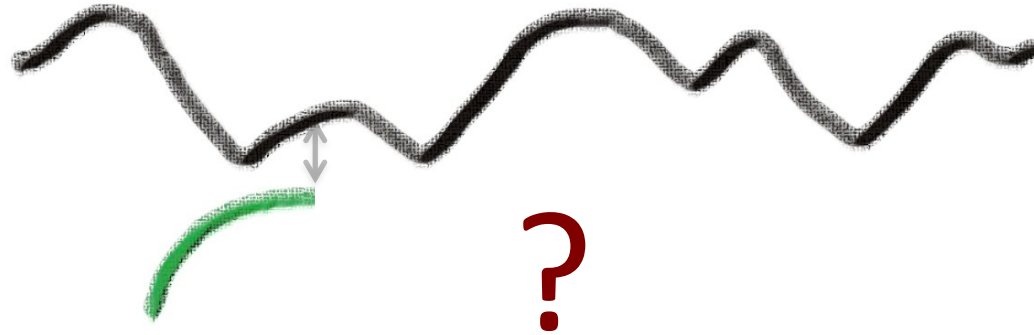
# Distance from the community



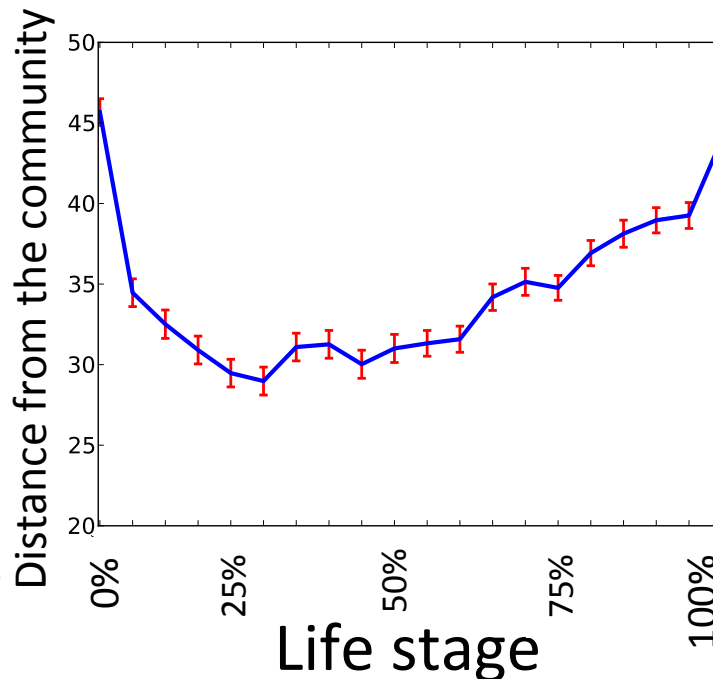
Stage 1:  
user **assimilates**  
the language of  
the community



# Distance from the community



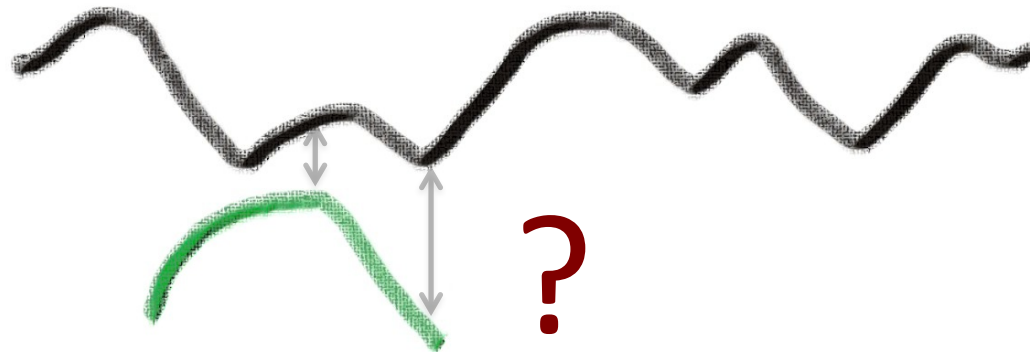
Stage 1:  
user **assimilates**  
the language of  
the community



Stage 2:  
User's language  
**distances** itself  
from that  
of the community

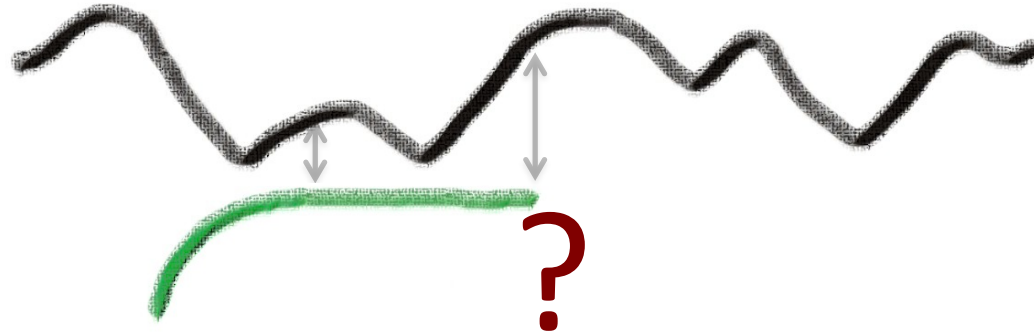


# Distance from the community



Hypothesis 1: User moves away from the community by **using innovative language**

# Distance from the community



Hypothesis 1: User moves away from the community by using innovative language

Hypothesis 2: User **stops adapting** and gets out of tune with the changing community

# User language stability

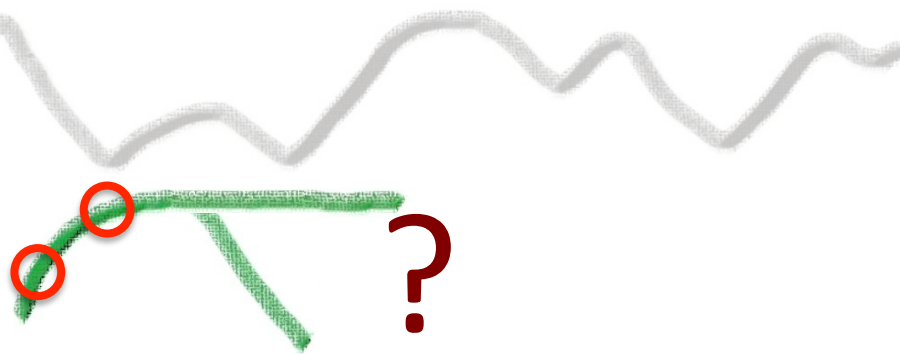


Hypothesis 1: User moves away from the community by **using innovative language**

Hypothesis 2: User **stops adapting** and gets out of tune with the changing community

# User language stability

Compare user language  
with her past language



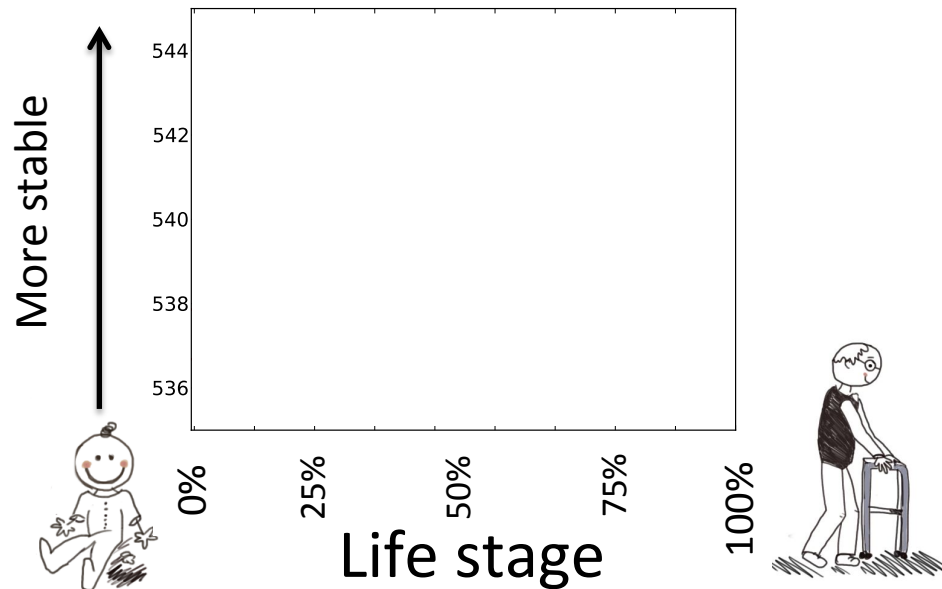
Hypothesis 1: User moves away from the community by  
**using innovative language**

Hypothesis 2: User **stops adapting** and gets out of tune with  
the changing community



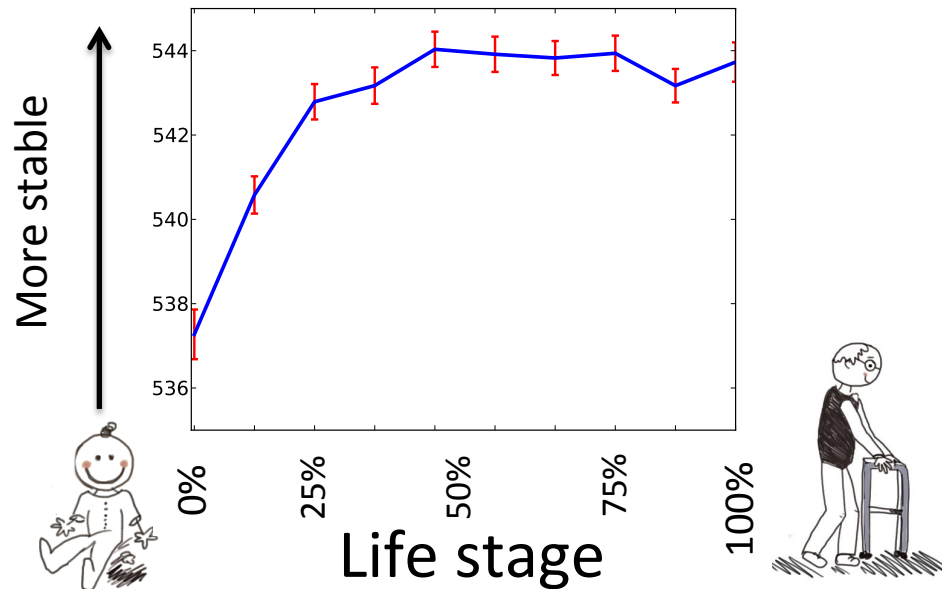
# User language stability

Compare user language with her past language



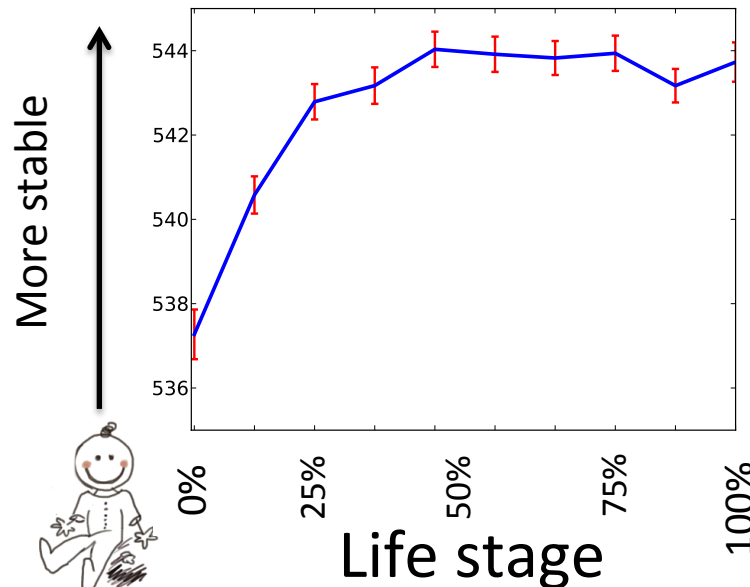
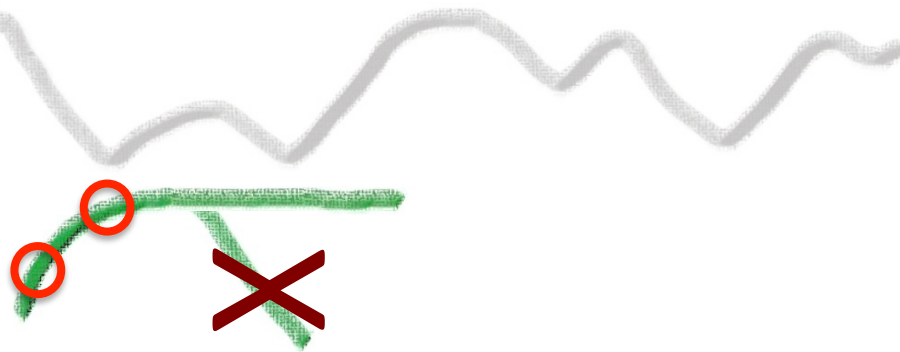
# User language stability

Compare user language with her past language



# User language stability

Compare user language with her past language



Confirms Hypothesis 2:  
before abandoning,  
users **stop adapting**



# Adoption of lexical innovation

# Adoption of lexical innovation

## Lexical innovation:

- new word that get picked up by the community
- about 100 lexical innovations each month

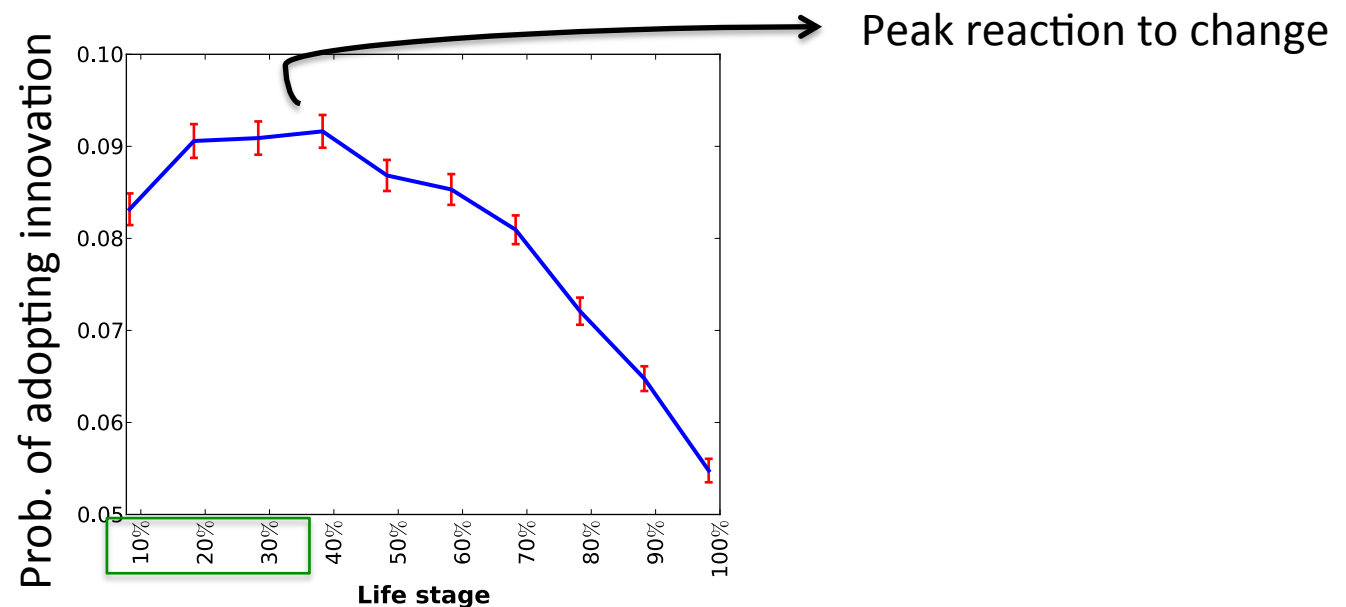
Type	Examples
Conventions	S[mell], M[outhfeel], FLAVOR
Descriptive	sandalwood, gummy, rubbery
Other	verdict, mysterious, nothingness

- user “adopts lexical innovation” if uses such a word in the 3 months after its introduction

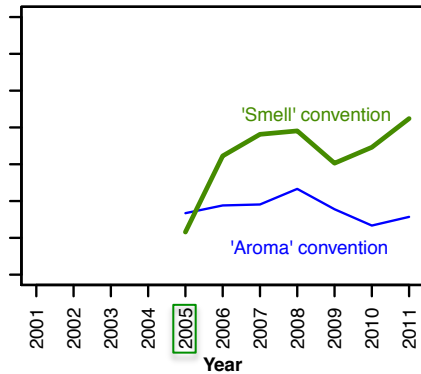
# Adoption of lexical innovation

## Lexical innovation:

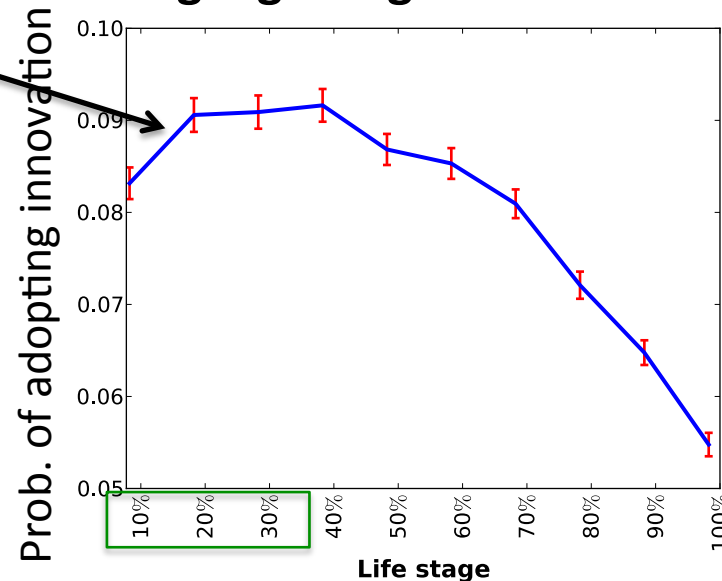
- new word that get picked up by the community



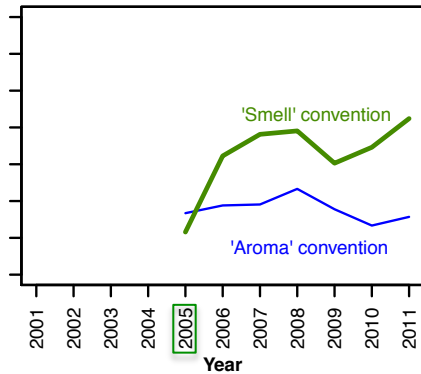
# Puzzle answer



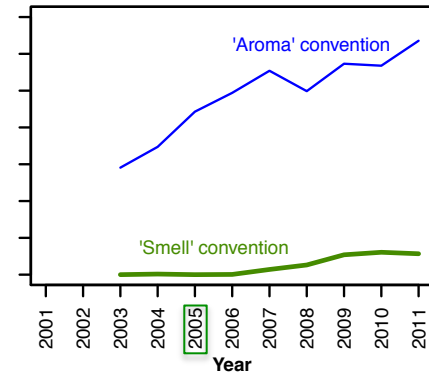
Users joining in 2005  
still in their **flexible-language stage**



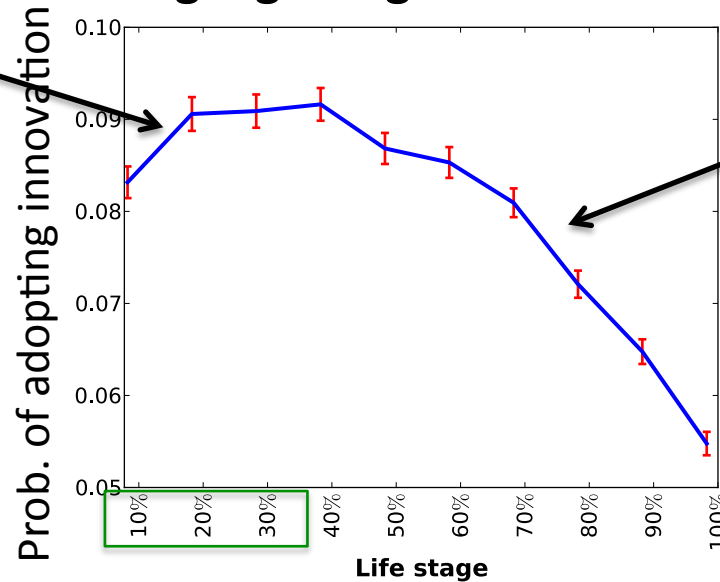
# Puzzle answer



Users joining in 2005  
still in their **flexible-language stage**



Users joining in 2003  
in the **rigid-language phase**





# User lifecycle (summary)

## Online linguistic lifecycle

0% User joins the community

### **Stage 1: adaptation to community norms**

30% Peak receptivity to community norms

### **Stage 2: linguistic patterns rigidify**

100% User abandons the community



# User lifecycle (summary)

## Online linguistic lifecycle

0% User joins the community

**Stage 1: adaptation to community norms**

30% Peak receptivity to community norms

**Stage 2: linguistic patterns rigidify**

100% User abandons the community

## Offline linguistic lifecycle [Labov, 1966]

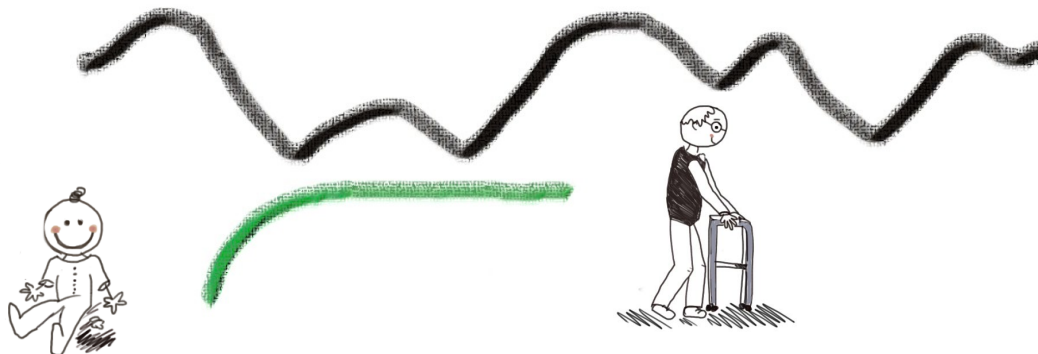
Birth Individual joins the community

**Stage 1: linguistic assimilation**

17 years Peak receptivity to community norms

**Stage 2: “adult language stability”**

Individual leaves the community



# User lifecycle (summary)

## Online linguistic lifecycle

0% User joins the community

### Stage 1: adaptation to community norms

30% Peak receptivity to community norms

### Stage 2: linguistic patterns rigidify

100% User abandons the community

## Offline linguistic lifecycle [Labov, 1966]

Birth Individual joins the community

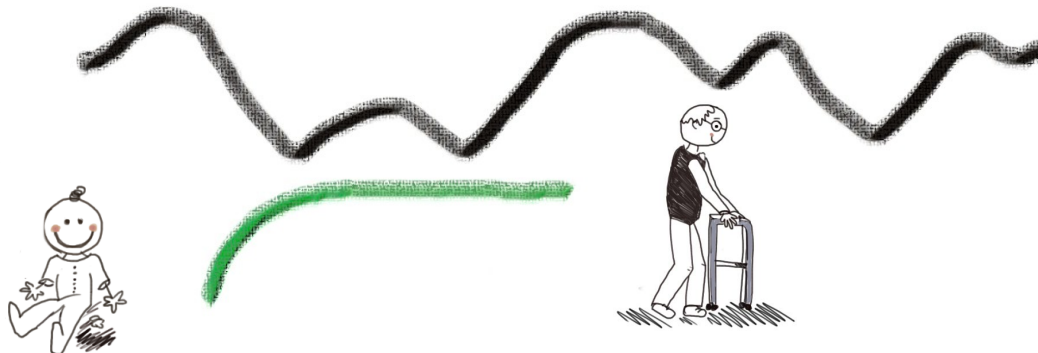
### Stage 1: linguistic assimilation

17 years Peak receptivity to community norms

### Stage 2: “adult language stability”

Individual leaves the community

Absolute time-frame,  
assumed biological effect



# User lifecycle (summary)

## Online linguistic lifecycle

0% User joins the community

### Stage 1: adaptation to community norms

**30%** Peak receptivity to community norms

### Stage 2: linguistic patterns rigidify

100% User abandons the community

## Offline linguistic lifecycle [Labov, 1966]

Birth Individual joins the community

### Stage 1: linguistic assimilation

**17 years** Peak receptivity to community norms

### Stage 2: “adult language stability”

Individual leaves the community

Relative time-frame,  
suggesting social effect

Absolute time-frame,  
assumed biological effect

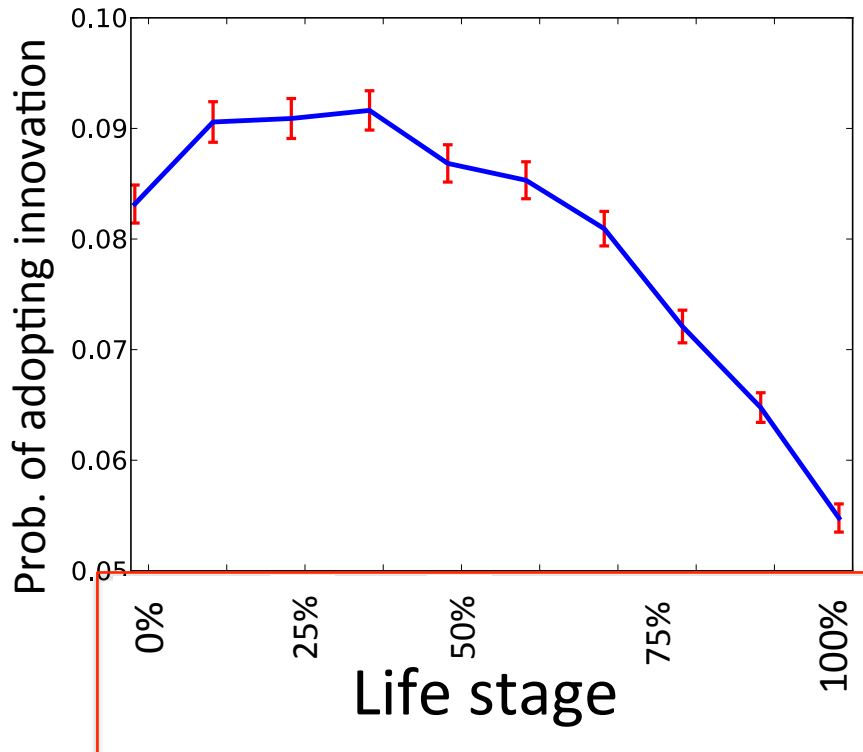


# Elastic lifecycle

Lifecycle **stretches** according to the user's ultimate lifespan

# Elastic lifecycle

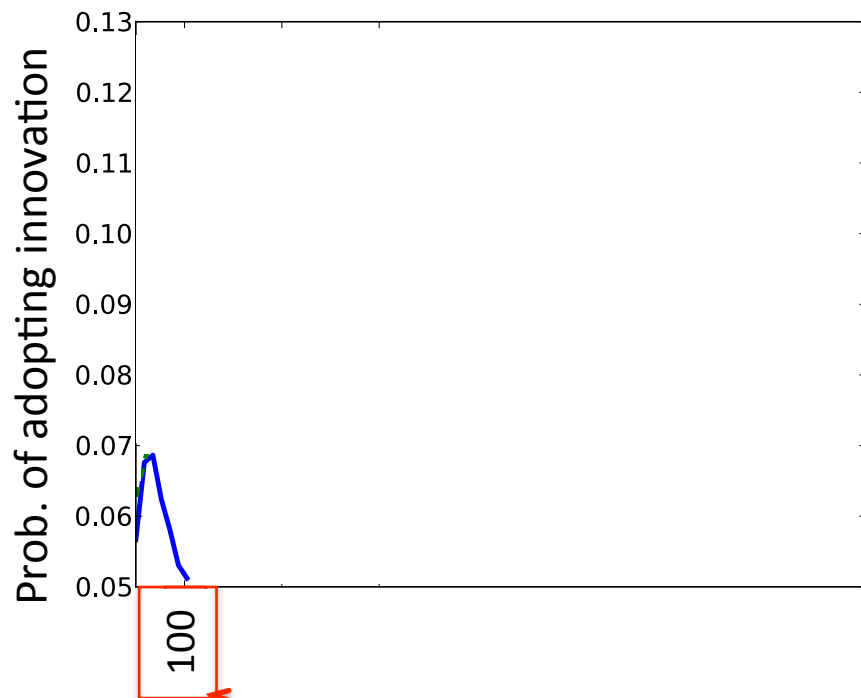
Lifecycle **stretches** according to the user's ultimate lifespan



Users with vastly different lifespans

# Elastic lifecycle

Lifecycle **stretches** according to the user's ultimate lifespan

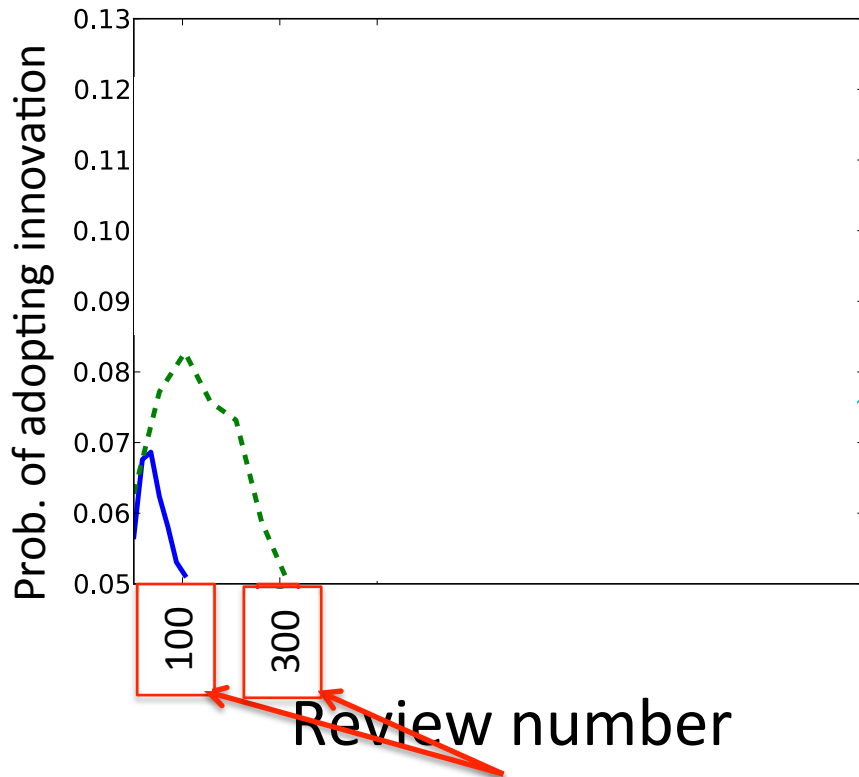


Review number

Break down by user ultimate lifespan

# Elastic lifecycle

Lifecycle **stretches** according to the user's ultimate lifespan

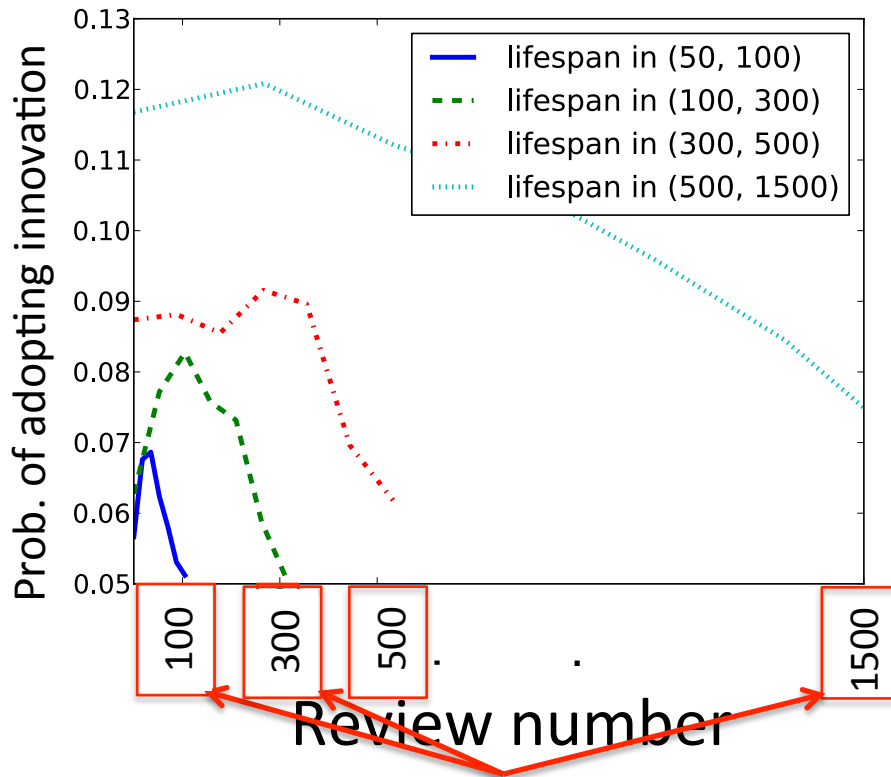


Break down by user ultimate lifespan



# Elastic lifecycle

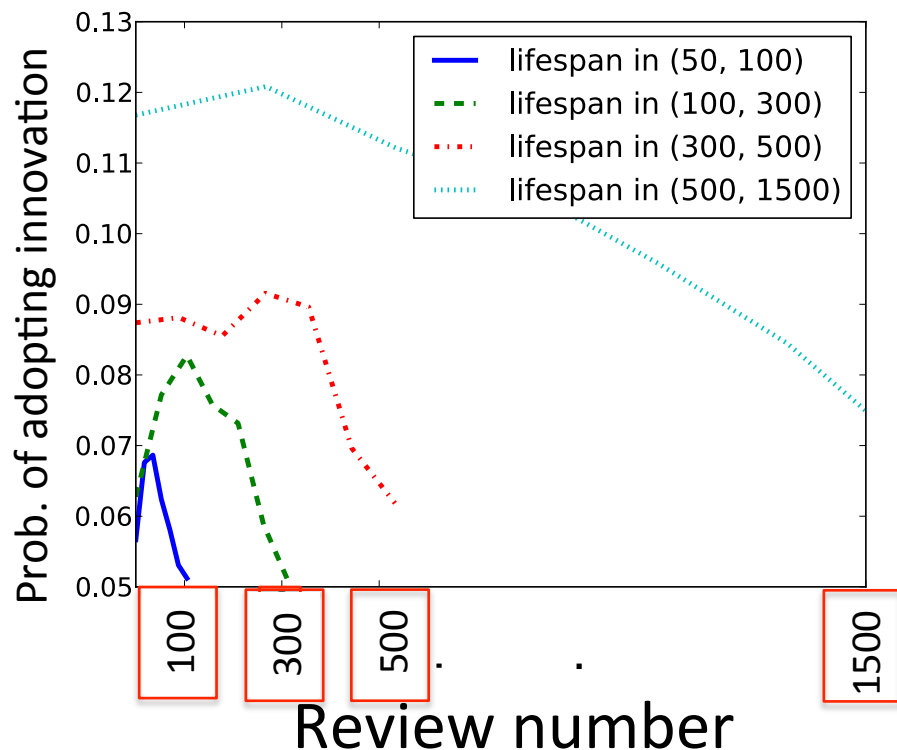
Lifecycle **stretches** according to the user's ultimate lifespan



Break down by user ultimate lifespan

# Elastic lifecycle

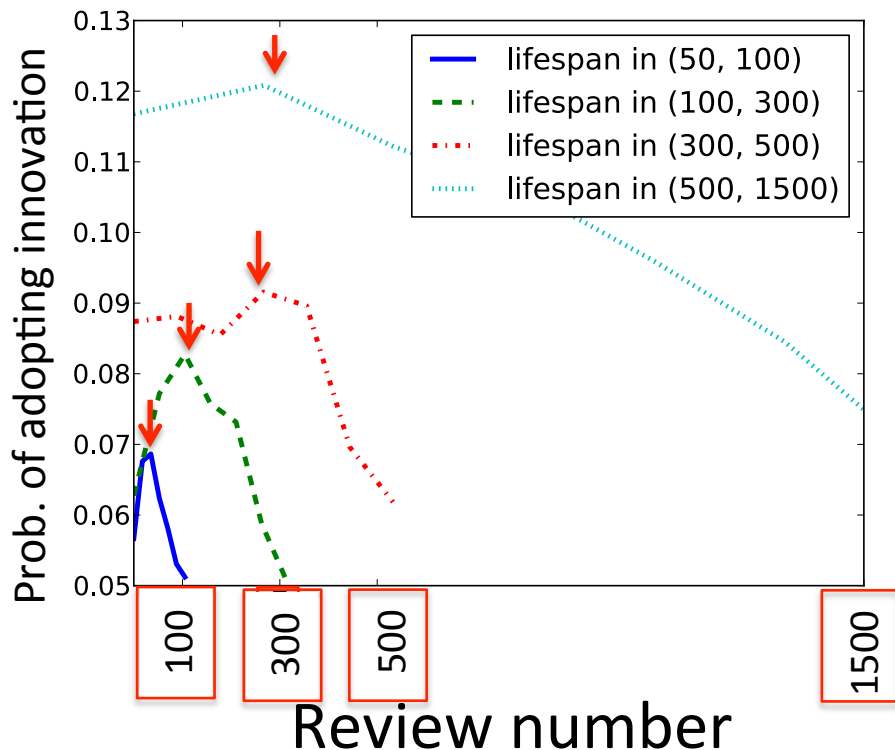
Lifecycle **stretches** according to the user's ultimate lifespan



→ **Similar lifecycle in spite of vastly different lifespans**  
“All users die old”

# Elastic lifecycle

Lifecycle **stretches** according to the user's ultimate lifespan



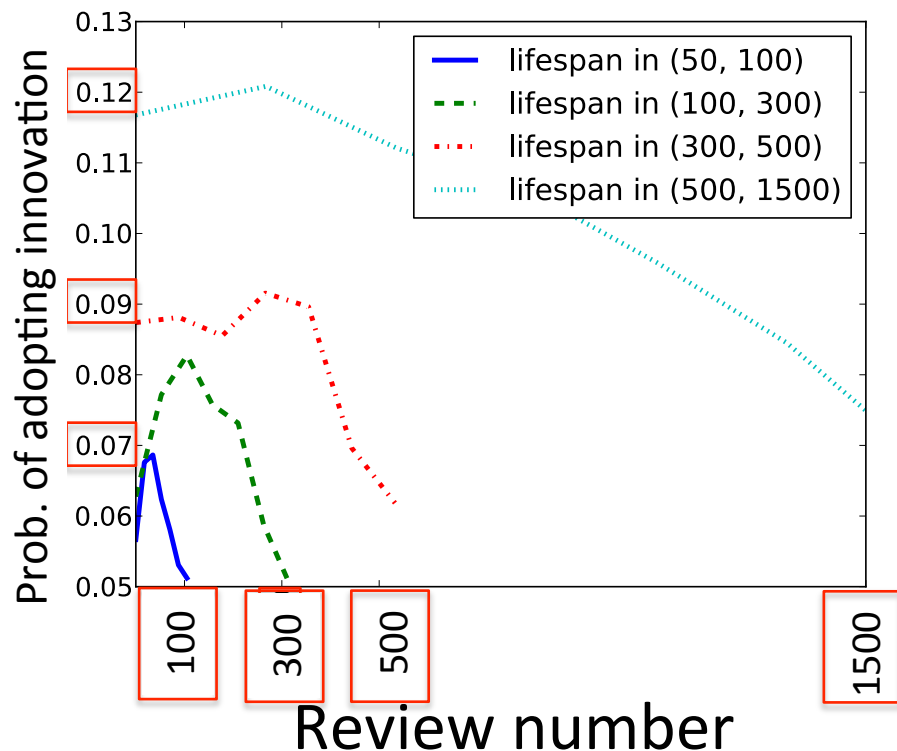
→ **Similar lifecycle in spite of vastly different lifespans**

“All users die old”

→ **End of Stage 1 is a function of the ultimate lifespan of the user**  
(not tied to an absolute timeframe  
e.g., 60 reviews or 1 year)

# Elastic lifecycle

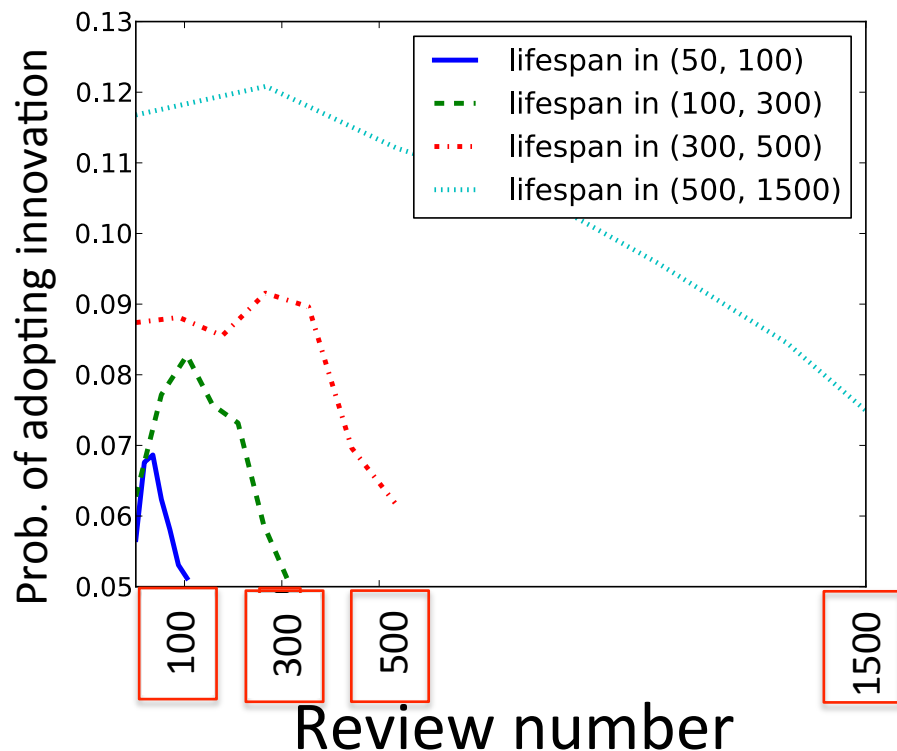
Lifecycle **stretches** according to the user's ultimate lifespan



- **Similar lifecycle in spite of vastly different lifespans**  
“All users die old”
- **End of Stage 1 is a function of the ultimate lifespan of the user**  
(not tied to an absolute timeframe  
e.g., 60 reviews or 1 year)
- **Level of receptivity is correlated with the ultimate lifespan of the user**

# Elastic lifecycle

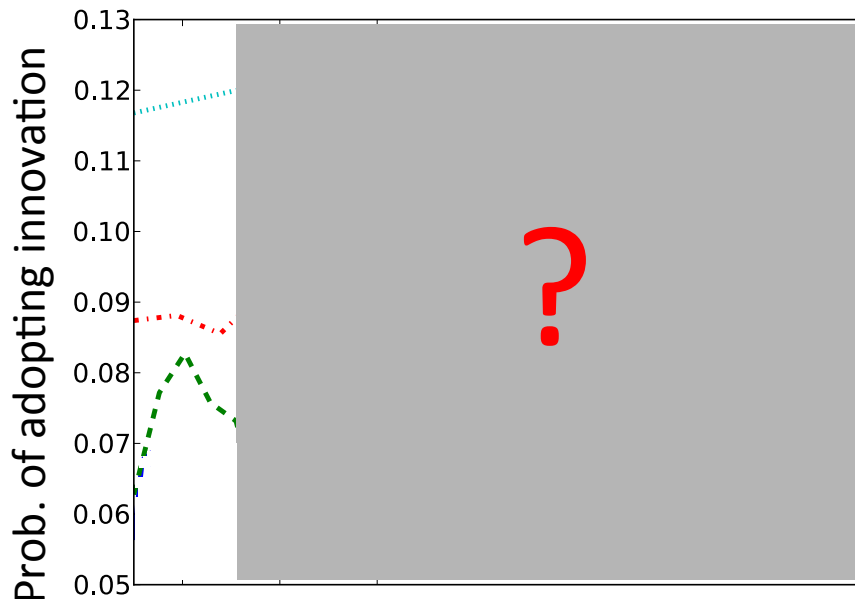
Lifecycle **stretches** according to the user's ultimate lifespan



- Similar lifecycle in spite of vastly **different lifespans**  
“All users die old”
- End of Stage 1 is a function of the **ultimate lifespan** of the user  
(not tied to an absolute timeframe  
e.g., 60 reviews or 1 year)
- Level of receptivity is correlated with the **ultimate lifespan** of the user

# Elastic lifecycle

Lifecycle **stretches** according to the user's ultimate lifespan



- Similar lifecycle in spite of vastly **different lifespans**  
“All users die old”
- End of Stage 1 is a function of the **ultimate lifespan** of the user  
(not tied to an absolute timeframe  
e.g., 60 reviews or 1 year)
- Level of receptivity is correlated  
with the **ultimate lifespan** of the user

**Predict ultimate lifespan**

# Predicting user lifespan

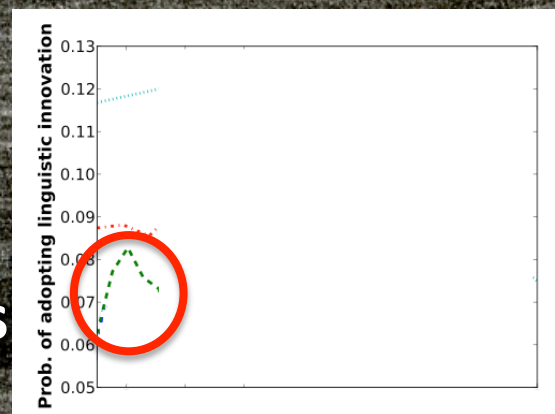
**Task: Given the first 20 posts,  
will the user abandon the community soon?**

# Predicting user lifespan

Task: Given the first 20 posts,  
will the user abandon the community soon?

**Linguistic change features:**

**distance from the community**  
**language stability**  
**adoption of lexical innovations**



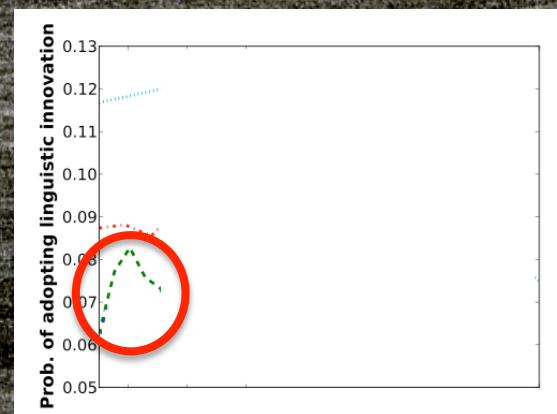


# Predicting user lifespan

Task: Given the first 20 posts,  
will the user abandon the community soon?

Linguistic change features:

- distance from the community
- language stability
- adoption of lexical innovations



Baselines:

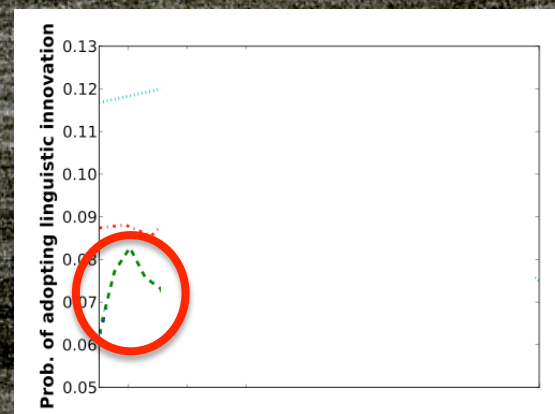
- post frequency ← previous work on churn prediction  
[Dror et al. 2012, Yang et al. 2010]
- post month ← accounts for community-wide changes

# Predicting user lifespan

Task: Given the first 20 posts,  
will the user abandon the community soon?

Linguistic change features:

- distance from the community
- language stability
- adoption of lexical innovations



Baselines:

- post frequency ← previous work on churn prediction [Dror et al. 2012, Yang et al. 2010]
- post month ← accounts for community-wide changes

**Logistic regression: One community for development, the other for test**

# Predicting user lifespan

**Results: Up to 12% absolute (40% relative) improvement**

<b>Features</b>	<b>F1</b>
<b>Baseline</b>	<b>30.5</b>
<b>+ Distance from the community</b>	<b>37.4</b>
<b>+ Language stability</b>	<b>38.0</b>
<b>+ Adoption of lexical innovation</b>	<b>40.9</b>
<b>+ First person singular pronouns</b>	<b>41.2</b>
<b>+ Number of words</b>	<b>42.9</b>

# Conclusions

- framework for tracking linguistic change
- revealed an elastic two-stage lifecycle
- exploited for predicting user abandonment
- co-evolution of users and their communities

# Thank you!



Data, slides, and more available at:  
[www.mpi-sws.org/~cristian](http://www.mpi-sws.org/~cristian)