

Análisis sociolingüístico de Twitter en las sociedades multilingües

اللغات

微博在多

Soziolinguistische Analyse von Twitter in mehrsprachigen Gesellschaften

다언어사회

Analyse sociolinguistique de Twitter dans les sociétés multilingues

多言語社

Sosiolinguististen analyysi Twitter fin Monikielinen Societies

Sociolinguistic Analysis of Twitter in Multilingual Societies

Suin Kim

Ingmar Weber, Li Wei, Alice Oh

School of Computing, KAIST

Birkbeck, University of London

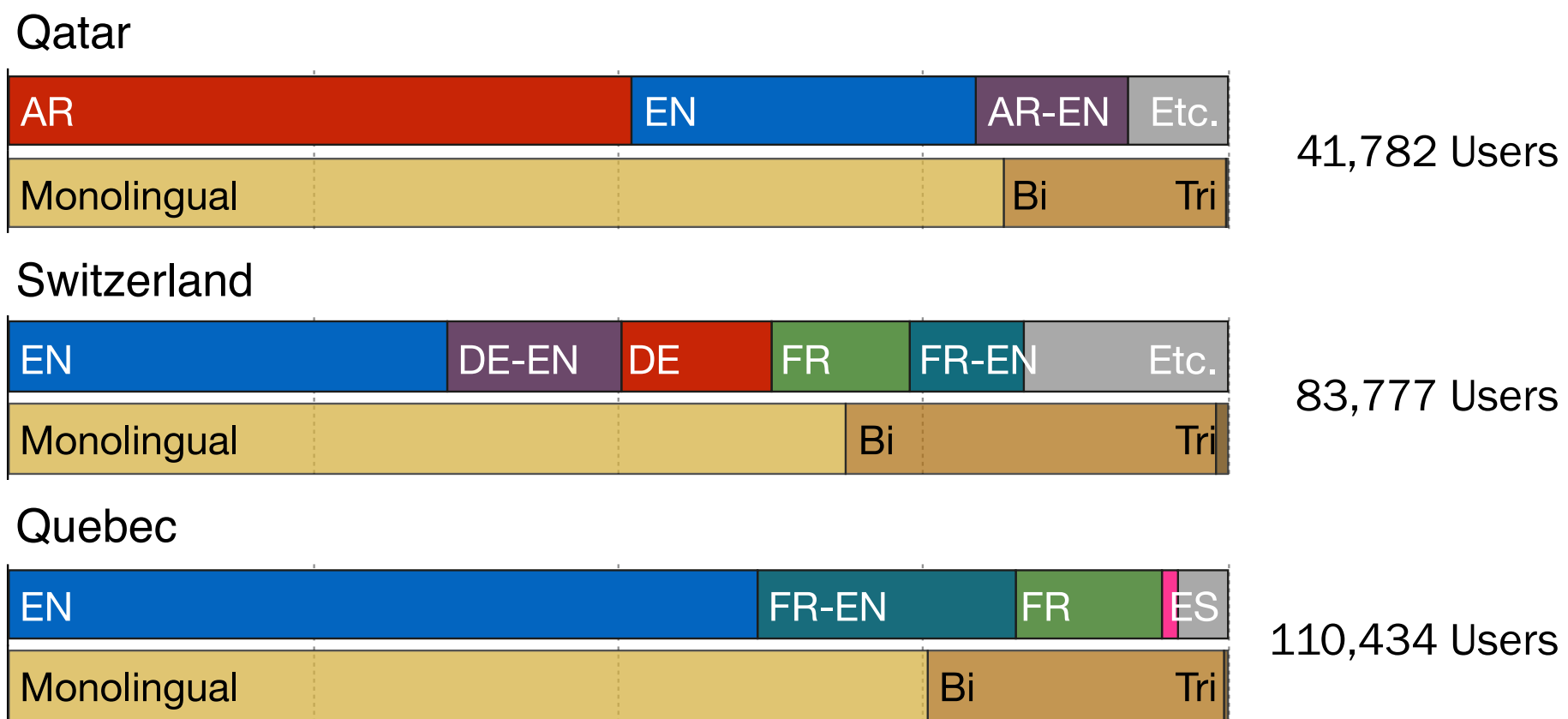
Qatar Computing Research Institute

Modern Society is Multilingual

- In the multilingual societies,
 - The language we speak is an integral part of our culture
 - A spoken language can be a political statement with wide-ranging implications
 - The dynamics of language choice could explain the prosperity or demise of local languages
- We study the phenomena of multilingual societies and the role that bilinguals play in them, questioning
 1. **Potential segregation:** Do monolinguals cluster together?
 2. **Social status:** Is there a “High Class” language?
 3. **Social capital:** Do bilinguals act as a bridge? If so, how?
 4. **Language preservation:** Will language converge to single lingua franca?
 5. **Code switching:** Do bilinguals use different language for different stories?

Twitter Data

- We captured users with geographic information in their profile
- We crawled all available tweets and followings
- We detected languages of tweet, classifying each user into language categories



Monolinguals Cluster Together

1. Potential Segregation

- *Do monolinguals cluster together?*

2. Social Status

- *Is there a “High Class” language?*

3. Social Capital

- *Do bilinguals act as a bridge? If so, how?*

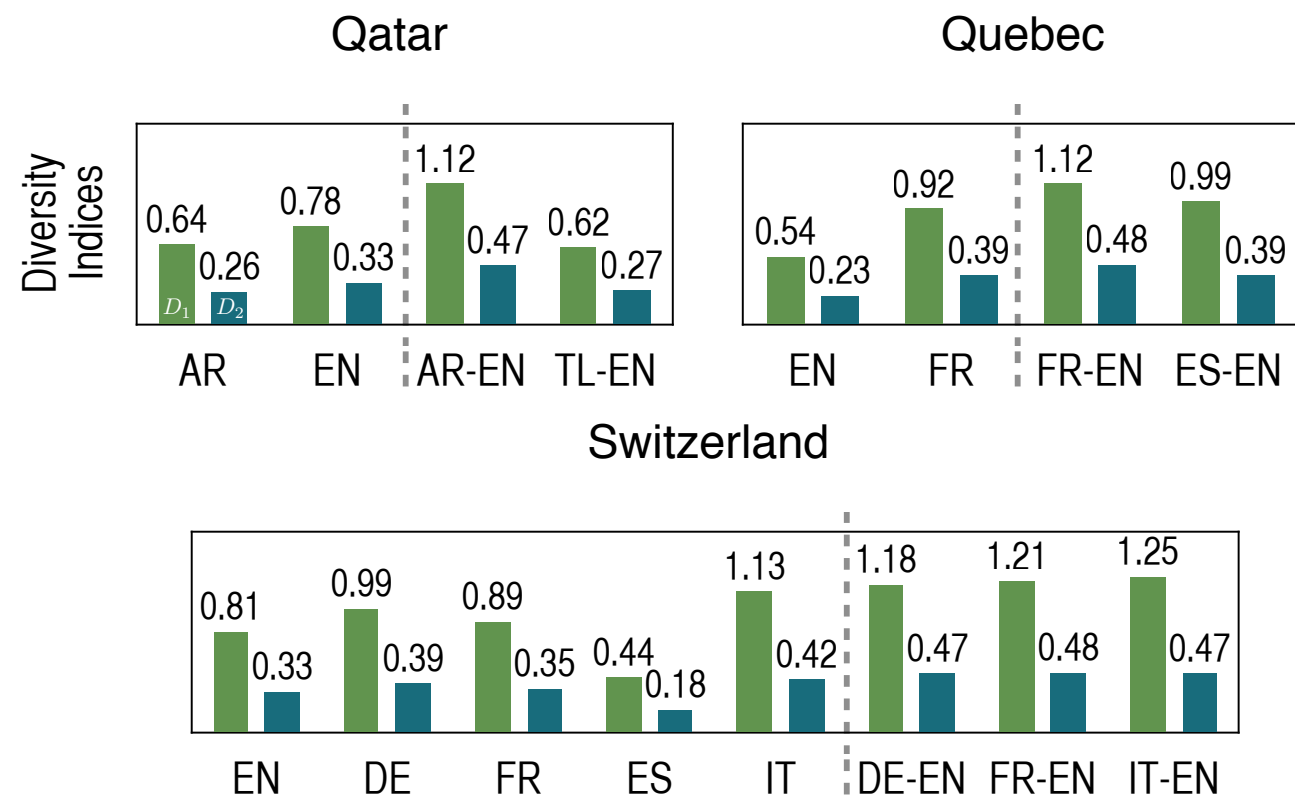
4. Language Preservation

- *Will language converge to single lingua franca?*

5. Code Switching

- *Do bilinguals use different language for different stories?*

- We analyzed network of followings between language groups
- **Diversity index:** How users make diverse connection in language speaking groups?



Monolinguals have lower diversity index
Bilinguals connect to more diverse users

Monolinguals Cluster Together

1. Potential Segregation

- *Do monolinguals cluster together?*

2. Social Status

- *Is there a “High Class” language?*

3. Social Capital

- *Do bilinguals act as a bridge? If so, how?*

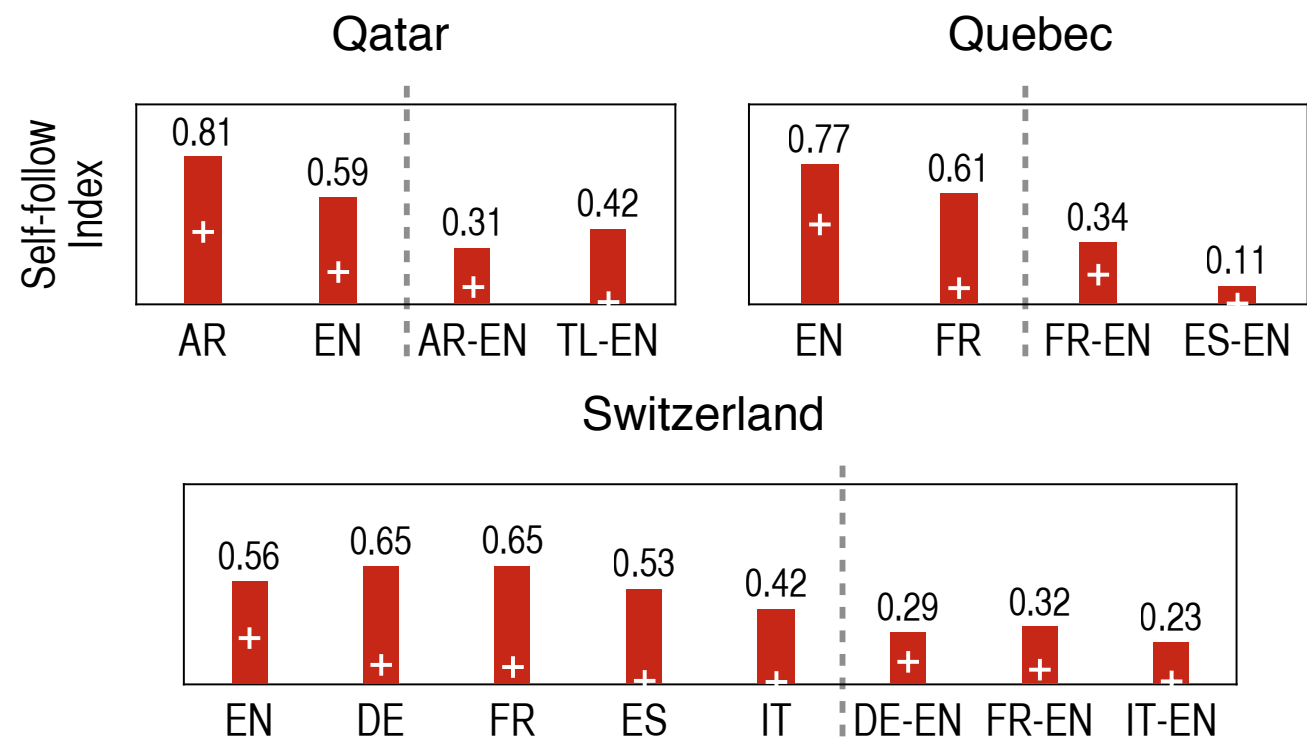
4. Language Preservation

- *Will language converge to single lingua franca?*

5. Code Switching

- *Do bilinguals use different language for different stories?*

- **Self-follow index:** How users tend to make homogeneous connections?



Monolinguals have lower diversity index
and higher self-follow index

Users of Local Languages Exert Higher Influence

1. Potential Segregation

- *Do monolinguals cluster together?*

2. Social Status

- *Is there a “High Class” language?*

3. Social Capital

- *Do bilinguals act as a bridge? If so, how?*

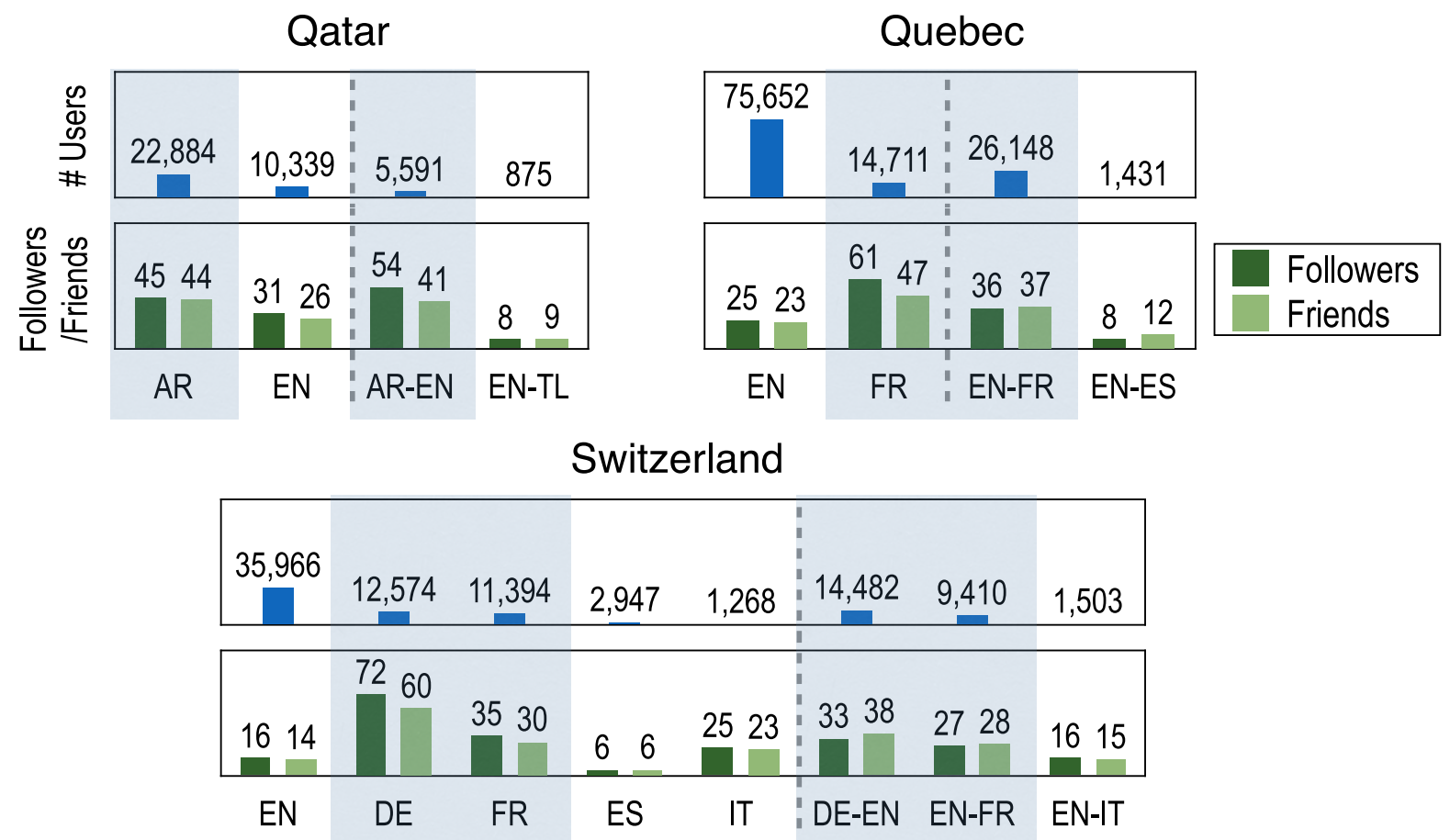
4. Language Preservation

- *Will language converge to single lingua franca?*

5. Code Switching

- *Do bilinguals use different language for different stories?*

- Studies have shown that tweets from a user with a high in-degree are more likely to be retweeted



Users of **local language** have higher followers than others even when there are more English monolinguals

Bilinguals Act as a Bridge

1. Potential Segregation

- *Do monolinguals cluster together?*

2. Social Status

- *Is there a “High Class” language?*

3. Social Capital

- *Do bilinguals act as a bridge? If so, how?*

4. Language Preservation

- *Will language converge to single lingua franca?*

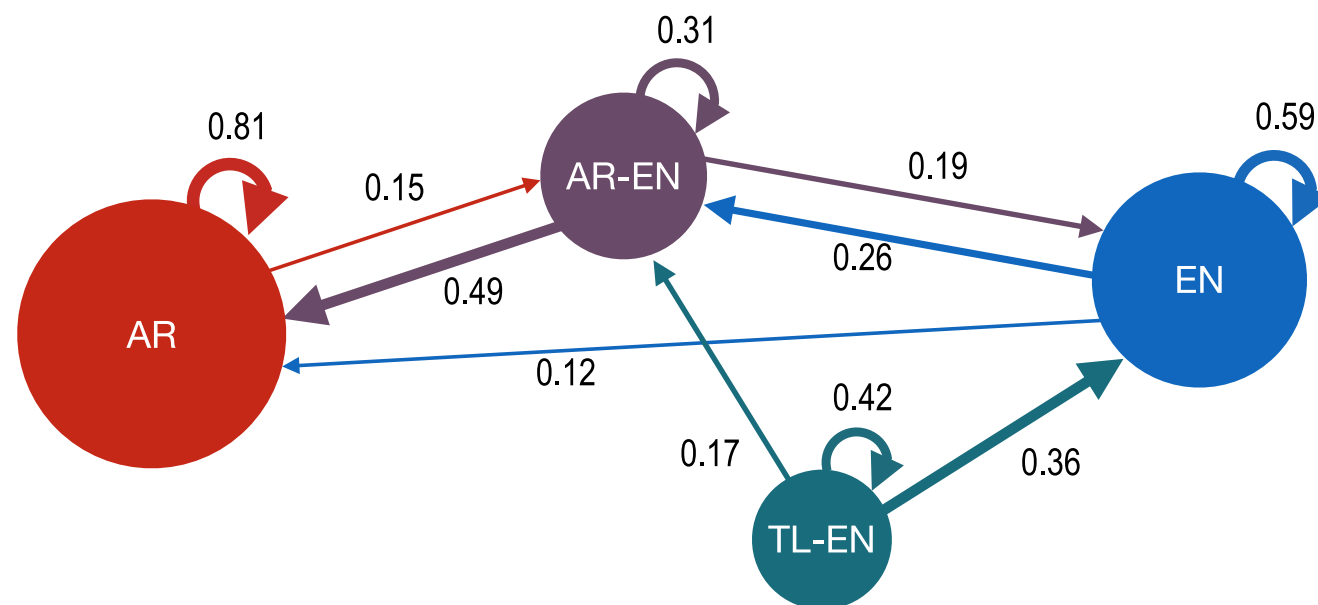
5. Code Switching

- *Do bilinguals use different language for different stories?*

Following Network (Qatar)

Per-user averaged ratio of following

AR-EN bilinguals bridge between AR and EN speakers



EN connects
AR and
TL speakers

*Following ratio < 0.10 not shown

Bilinguals Act as a Bridge

1. Potential Segregation

- *Do monolinguals cluster together?*

2. Social Status

- *Is there a “High Class” language?*

3. Social Capital

- *Do bilinguals act as a bridge? If so, how?*

4. Language Preservation

- *Will language converge to single lingua franca?*

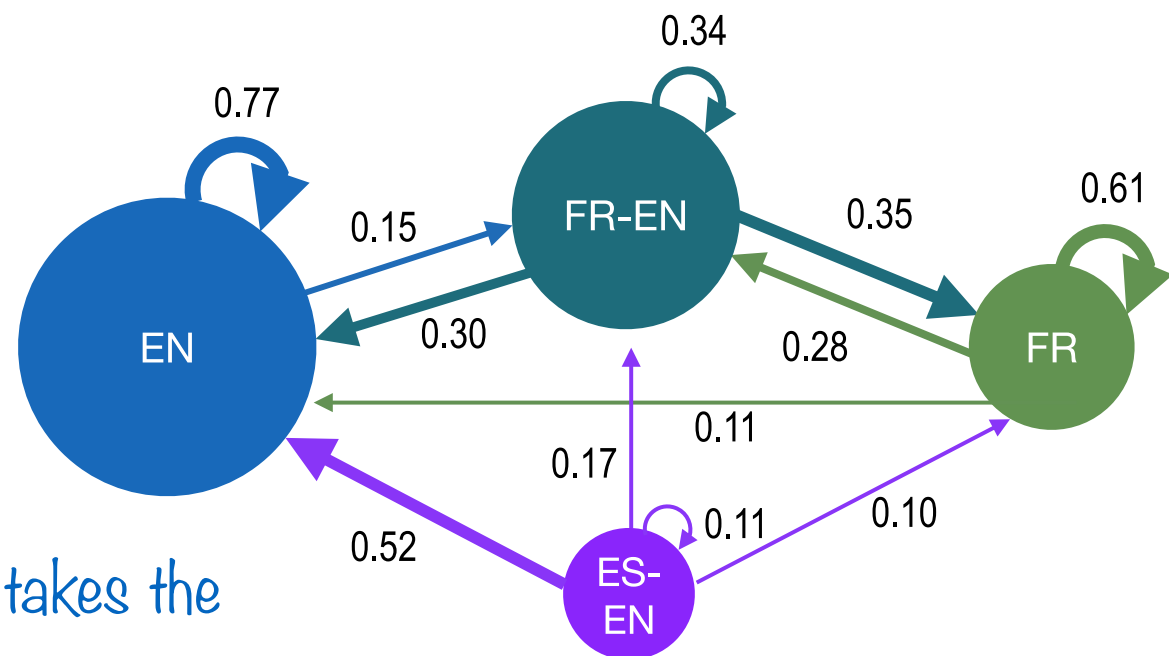
5. Code Switching

- *Do bilinguals use different language for different stories?*

Following Network (Quebec)

Per-user averaged ratio of following

FR-EN bilinguals bridge FR and EN speakers



Self-follow takes the highest proportion for monolinguals

Bilinguals Act as a Bridge

1. Potential Segregation

- *Do monolinguals cluster together?*

2. Social Status

- *Is there a “High Class” language?*

3. Social Capital

- *Do bilinguals act as a bridge? If so, how?*

4. Language Preservation

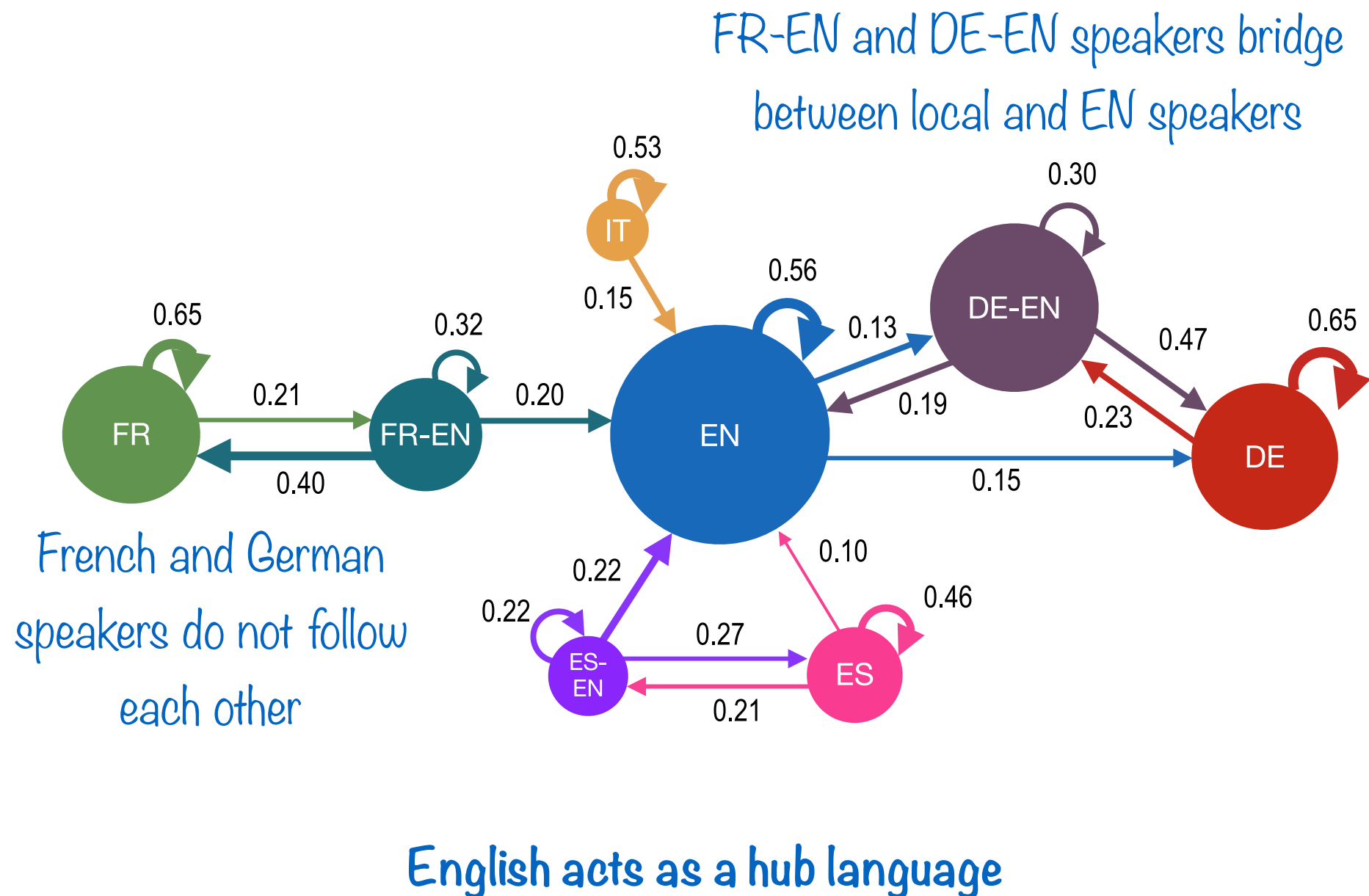
- *Will language converge to single lingua franca?*

5. Code Switching

- *Do bilinguals use different language for different stories?*

Following Network (Switzerland)

Per-user averaged ratio of following



English as Lingua Franca in Multilingual Twitter

1. Potential Segregation

- *Do monolinguals cluster together?*

2. Social Status

- *Is there a “High Class” language?*

3. Social Capital

- *Do bilinguals act as a bridge? If so, how?*

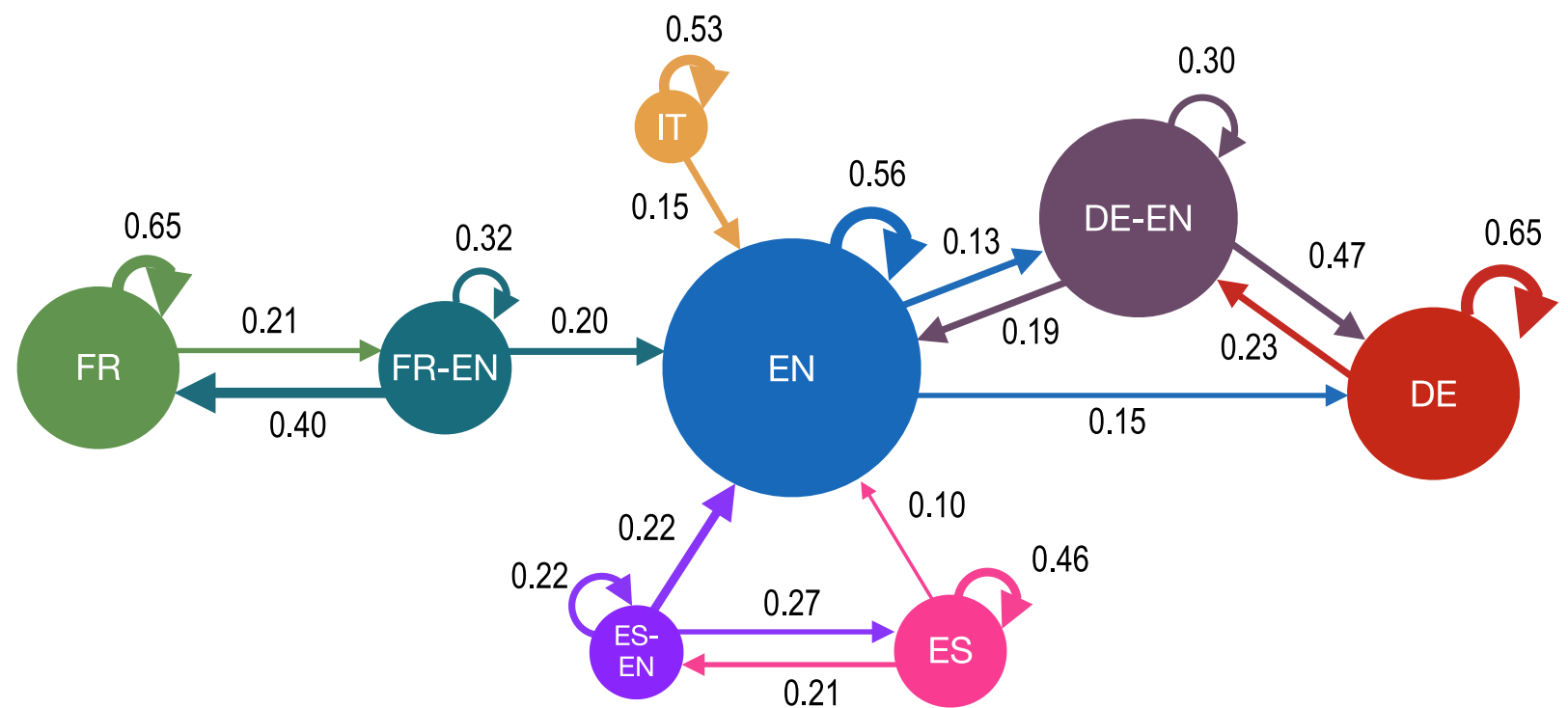
4. Language Preservation

- *Will language converge to single lingua franca?*

5. Code Switching

- *Do bilinguals use different language for different stories?*

- English acts as a hub language
- Bilingual group bridges two monolingual groups
- In-group following takes the largest proportion for monolingual groups
- Monolingual users do not follow other monolingual users



Language Usage and Audience

1. Potential Segregation

- *Do monolinguals cluster together?*

2. Social Status

- *Is there a “High Class” language?*

3. Social Capital

- *Do bilinguals act as a bridge? If so, how?*

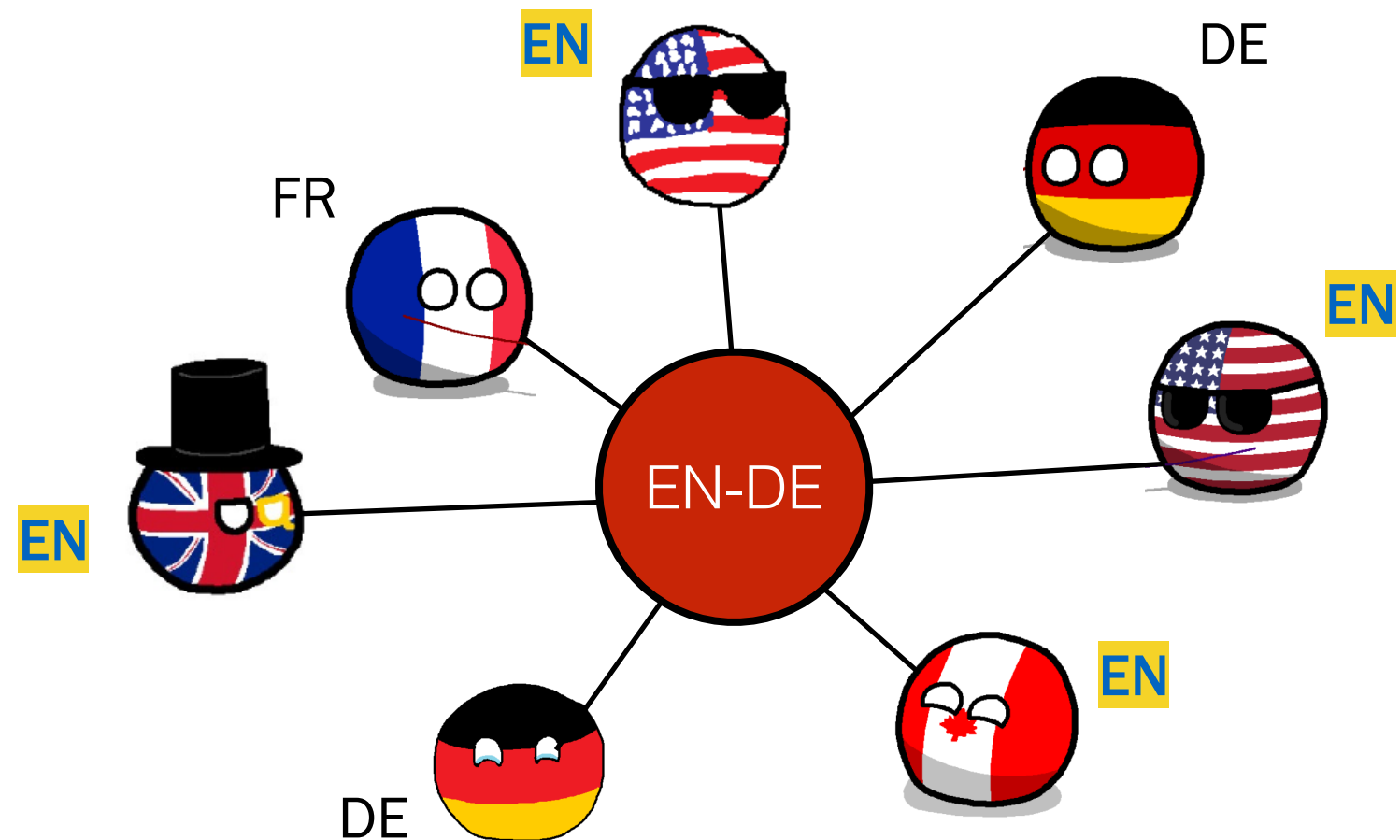
4. Language Preservation

- *Will language converge to single lingua franca?*

5. Code Switching

- *Do bilinguals use different language for different stories?*

- How should multilingual users choose their language in Twitter?



A game-theoretic approach to “maximize the audience”
will force the user to tweet in English

Are Minor Languages Prone to “Die Out”?

1. Potential Segregation

- *Do monolinguals cluster together?*

2. Social Status

- *Is there a “High Class” language?*

3. Social Capital

- *Do bilinguals act as a bridge? If so, how?*

4. Language Preservation

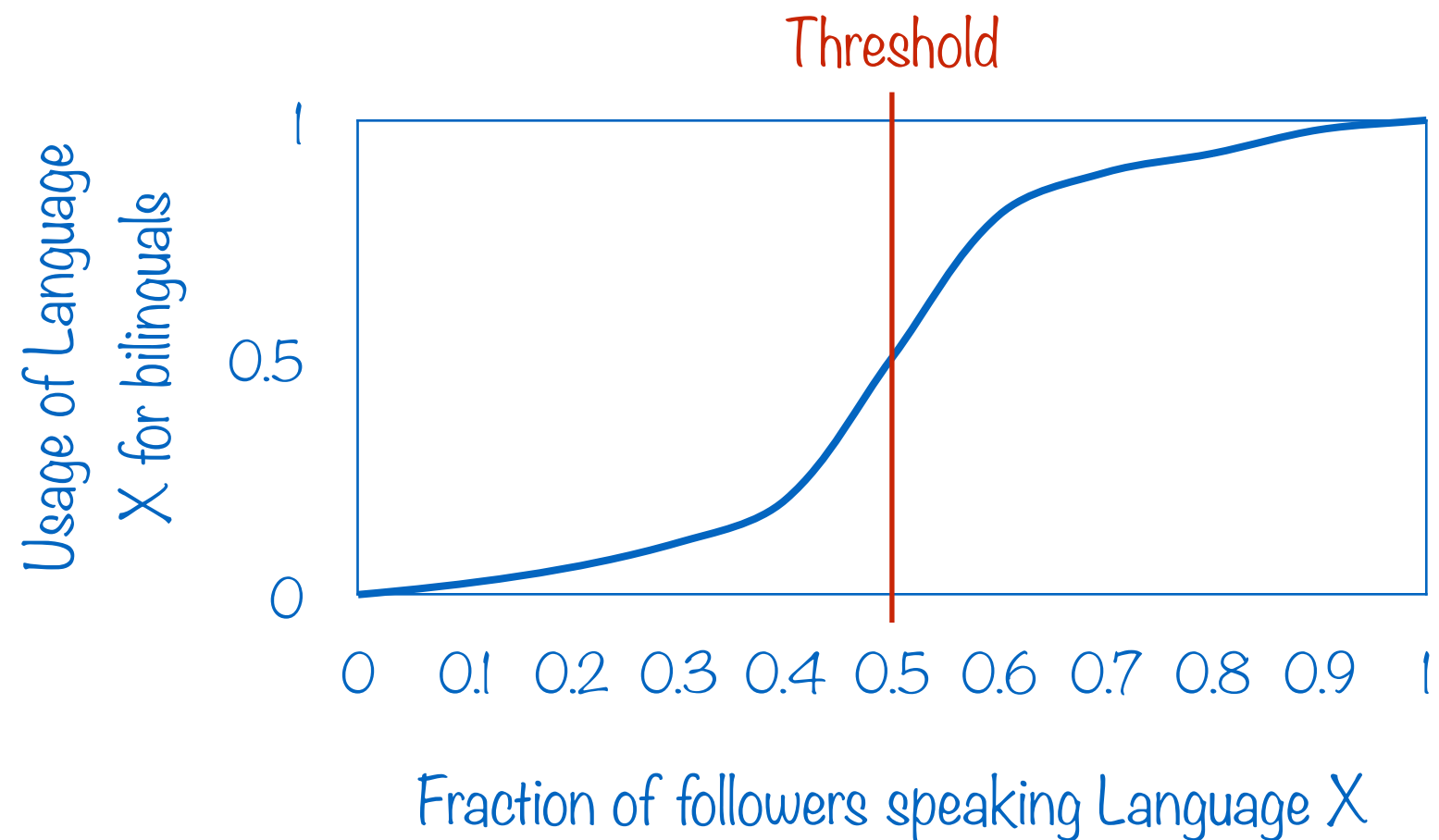
- *Will language converge to single lingua franca?*

5. Code Switching

- *Do bilinguals use different language for different stories?*

• Our Expectation

- Sudden leap of language usage after the threshold



- Will it lead to a global convergence to a single lingua franca?
- What is the **threshold** that users change their language?

Languages will not Converge

1. Potential Segregation

- *Do monolinguals cluster together?*

2. Social Status

- *Is there a “High Class” language?*

3. Social Capital

- *Do bilinguals act as a bridge? If so, how?*

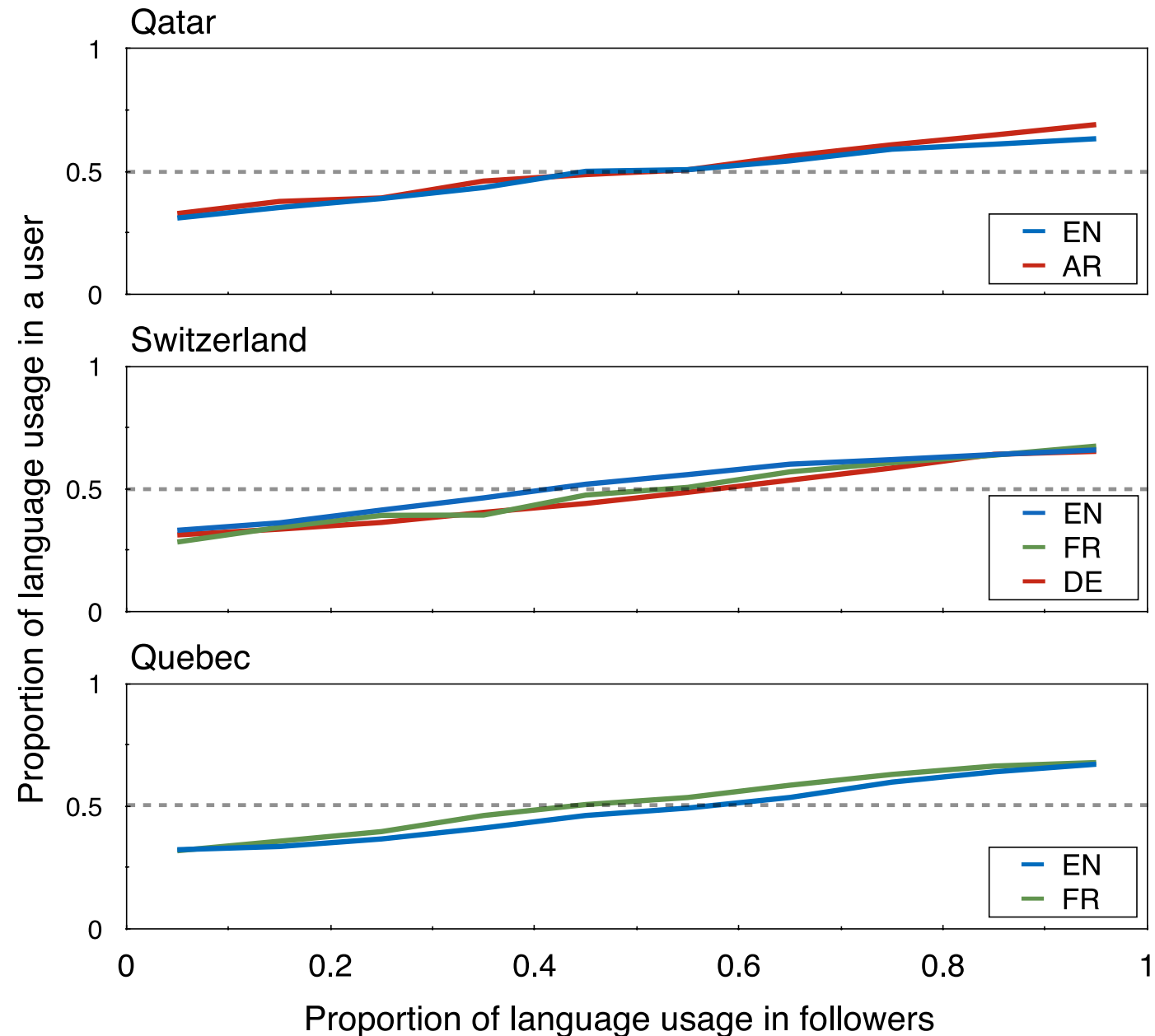
4. Language Preservation

- *Will language converge to single lingua franca?*

5. Code Switching

- *Do bilinguals use different language for different stories?*

Bilingual users “mimic” the language mix observed along their followers



A non-steep, centered language mix indicates the language convergence is not being accelerated

Bilinguals Use Different Language for Different Stories

1. Potential Segregation

- *Do monolinguals cluster together?*

2. Social Status

- *Is there a “High Class” language?*

3. Social Capital

- *Do bilinguals act as a bridge? If so, how?*

4. Language Preservation

- *Will language converge to single lingua franca?*

5. Code Switching

- *Do bilinguals use different language for different stories?*

- Use probabilistic topic model (LDA) on tweets from bilinguals (topics=10)
- Grouped tweets having hashtags with same meaning but different languages: (#switzerland—#suisse, #qatar—قطر#)

Bilingual users post
informational and political
tweets in local language

events, tourism, photography,
leisure-related tweets
in English for the
non-local audience

QATAR, AR-EN BILINGUAL		
#	Label	Top words
قطر#	development government emir/god gcc countries	national doha development vision government national qa foundation hamad bin god sheikh tamim emir doha egypt kuwait bahrain uae
#qatar	national day sports event photography recruitment	day national doha gcc about today doha volleyball football uae photo instagram katara instagood love job please send cv recruitment org
SWITZERLAND, EN-FR BILINGUAL		
#suisse	politics news/radio recruitment ski	more country politics no federal ch news tar info thank you radio job manager head senior engineer ski weather romandie snow rentals
#switzerland	scenery greeting party/club wine/fashion	lake sun sky sunset beautiful night love like time good show morning club dj party enjoy welcome house fashion valais basel wine beautiful

Summary

A large-scale computational analysis of language use and network characteristics in multilingual societies

1. Potential Segregation

- *Do monolinguals cluster together?*

2. Social Status

- *Is there a “High Class” language?*

3. Social Capital

- *Do bilinguals act as a bridge? If so, how?*

4. Language Preservation

- *Will language converge to single lingua franca?*

5. Code Switching

- *Do bilinguals use different language for different stories?*

1. **Monolinguals cluster together**, while bilinguals do not

2. Users speaking **local language** have higher influence

3. **Bilinguals bridge between monolinguals** with English as a hub

4. Language-mixing ratio of bilingual users **mirrors the mix of their followership**

5. Bilinguals express **informative/political/debatable topics in a local language**, while posting **event/tour/enjoyment topics on the English**

Thank You

Suin Kim
suin.kim@kaist.ac.kr
uilab.kaist.ac.kr/members/suinkim/