Sociolinguistic Analysis of Twitter in Multilingual Societies*

Suin Kim
KAIST
Republic of Korea
suin.kim@kaist.ac.kr

Ingmar Weber
Qatar Computing
Research Institute
iweber@qf.org.qa

Li Wei
Birkbeck,
University of London
li.wei@bbk.ac.uk

Alice Oh
KAIST
Republic of Korea
alice.oh@kaist.edu

ABSTRACT

In a multilingual society, language not only reflects culture and heritage, but also has implications for social status and the degree of integration in society. Different languages can be a barrier between monolingual communities, and the dynamics of language choice could explain the prosperity or demise of local languages in an international setting. We study this interplay of language and network structure in diverse, multi-lingual societies, using Twitter. In our analysis, we are particularly interested in the role of bilinguals. Concretely, we attempt to quantify the degree to which users are the "bridge-builders" between monolingual language groups, while monolingual users cluster together. Also, with the revalidation of English as a *lingua franca* on Twitter, we reveal users of the native non-English language have higher influence than English users, and the language convergence pattern is consistent across the regions. Furthermore, we explore for which topics these users prefer their native language rather than English. To the best of our knowledge, this is the largest sociolinguistic study in a network setting.

Categories and Subject Descriptors

J.4 [**Social and Behavioral Sciences**]: Sociology

Keywords

Multilingualism; Sociolinguistics; Topic Modeling; Social Media

1. INTRODUCTION

The language we speak is an integral part of our culture. We use it to communicate, to transmit facts and emotions, and to navigate the social environment surrounding us. In multilingual societies such as Canada or Switzerland, the spoken language can even be a political statement with wide-ranging implications. In some cases, governments subsidize programs to save a language from disappearing. At the individual level, people who are fortunate to be bilingual constantly make a choice in favor of or against one or the other language. Anecdotally, even after mastering a second language many people continue to count (and swear [14]) in their mother tongue. At the societal level, the question arises to which degree bilinguals are the "glue" that keeps multilingual societies together.

We study the phenomena of multilingual societies and the role that bilinguals play in them by using large amounts of Twitter data. Social media data has the fascinating component of also containing a *network*. These social links allow us to investigate the interaction between a user's language and their social surroundings. Understanding this interaction has a number of potential implications:

- *Preservation of a language.* Assuming that you are bilingual and that all of your friends understand English reasonably well, but not all understand your native language. Should you switch your language to maximize your audience size?
- *Social capital*[8] *and potential issues of segregation.* Is it possible to build social ties across language barriers? Which role do bilinguals play in this "bridge-building"?
- *Social status and language assimilation.* Eliza Doolittle in George Bernard Shaw's "Pygmalion"/"My Fair Lady" underwent a huge change in social status by learning a new language, though just a "high class dialect" in this case. Generally, are there elite languages in multi-lingual societies?
- *Language selection.* How do bilinguals choose one language over the other for a given topic? Do they prefer their mother tongue for issues "close to the heart"? Correspondingly, is the same topic discussed differently in different languages?

We explore these questions with large-scale Twitter data from several multilingual societies. We analyze the Twitter following behavior to uncover whether monolingual users form tightly connected clusters, what bridging roles multilingual users play, and which language groups show higher social status. We apply language processing to analyze the amount of language usage depending on the surrounding network, and probabilistic topic modeling to discover the differences in topics in different languages by multilingual users. Methodologically, we propose metrics for quantifying language use and network diversity in multilingual Twitter network, and we illustrate techniques from machine learning applied to multilingual tweets.

*This work was done while the first author was at Qatar Computing Research Institute.

This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definite version was published in HT'14. http://dx.doi.org/10.1145/2631775.2631824.

2. RELATED WORK

There is now widespread recognition among linguists that social media such as Twitter are highly multilingual and provide an immense volume of real-world language data. Several studies in sociolinguistics have explored Twitter and other online language [6, 9, 3, 22, 12, 34, 17, 2, 4], and in particular, social scientists have examined the strategic use of multilingualism on Twitter in recent political movements [30]. However, these studies are limited in scope, which require computational tools for a systematic analysis of multilingualism that involves both network analysis and language processing at a large scale.

More recently, sociolinguists, together with computer scientists, have tried to map out linguistic diversity through spatial and temporal analyses of multilingual Twitter. Studies by [5] and [24] have revealed the extensive use of a large number of different languages in Manchester, discovering Twitter users in Manchester are connected globally and use languages other than those recorded in the local census. Similarly, [25] used Twitter to detect the extent of multilingualism in London, which revealed that there are specific geographical concentrations of monolingual users of different languages. The processes of language shift, language attrition, language loss, language endangerment and language death were also investigated [19, 15, 18, 23, 31]. The related but different process is competition between different groups of language users [33, 29]. While groups with more socio-economic power often have a crucial impact on the spread of particular language, the size of the speaker group also plays a significant role. It has been shown that a single monolingual speaker of a particular language may hold the key to the survival of the language in the bilingual community, as the bilingual speakers try to accommodate the monolingual speaker [16, 20]. It then follows that relatively low number of language users could have a snowballing effect and prompt the majority to use a specific language in Twitter. A number of mathematical models for language competition have been proposed [1, 10, 28].

Social scientists, especially sociolinguists, have long been interested in the role language plays in the formation of social networks and in how structures of social networks impact on language practices [26, 13]. Relatively little is known about the role multilingualism plays in forming these networks and how the virtual networks impact on multilingual practices. While it is expected that speakers would identify themselves more easily with others who share the same languages, therefore forming language-specific clusters, it is not clear how monolinguals and bilinguals would pattern in relation to each other. Understanding the pattern of connections between monolingual and bilingual speakers would not only offer a new perspective on multilingualism on the social media, but also provide new insights into the societal structures and human relations in multilingual societies.

3. DATA COLLECTION AND LANGUAGE DETECTION

We collected Twitter data (recent tweets and friend/follower lists) from two countries (Qatar, Switzerland) and Quebec province in Canada. We identified Twitter accounts from two sources. First, we used Twitter streams provided by GNIP as part of a trial period. These streams comprise (i) about 28 hours of Firehose stream (ii) two weeks of decahose stream, both around June to August 2013. Users with at least one geo-tagged tweet from Qatar and Switzerland are considered as candidates. Another source of Twitter accounts is the location information from the public user profile using Followerwonk[1]. To capture users with only the city names, for each of the three countries we compiled lists of cities with more than 10^6 inhabitants, along with the names translated into multiple languages using Wikipedia entries. After identifying users from the regions of interest, we used the Twitter API to crawl all their *friends (= followings)* and *followers*, and up to 3,200 recent tweets for each user. Table 1 shows the statistics of the Twitter data. We ignored inactive users with fewer than 5 tweets. Then, we classified the language used in each tweet, which is not a trivial task because a large number of tweets contain very little information for language classification [24] and a single tweet can mix multiple languages. To optimize language classification accuracy against these challenges, we aggregated all tweets for each user into a document of tweets. After removing mentions, hashtags, and URLs, we used Compact Language Detector 2 [2] and detected the top three languages with their approximate percentages of the text bytes in the document. We define users as speaking a language if they have $\geq 15\%$ of text bytes written in the respective language. Figure 2 shows the distribution of language usage.

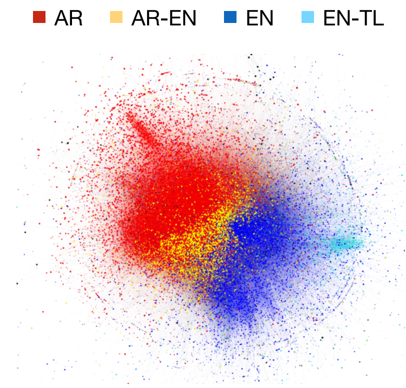

Figure 1: Visualization of Qatar Twitter network. Each node and edge represents user and followings in the Twitter networks. Each node is colored by the language usage from corresponding user's tweets. AR-EN Bilingual users are located between monolingual clusters.

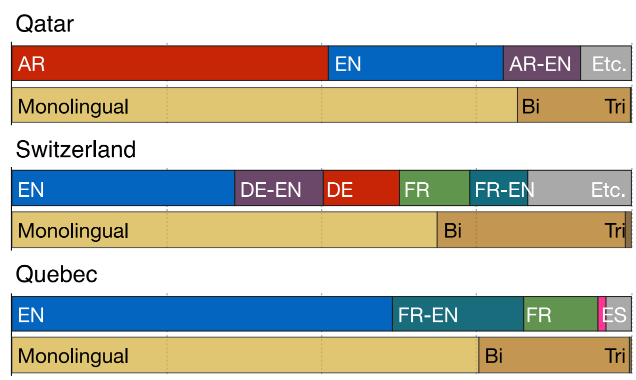

Figure 2: Language distribution of the Twitter users for each region. The upper bar illustrates the distribution of language usage, and the lower bar shows the distribution of mono-, bi-, and trilingual users. For all regions there are $< 1\%$ of trilingual users.

[1] https://followerwonk.com/bio
[2] https://code.google.com/p/cld2/

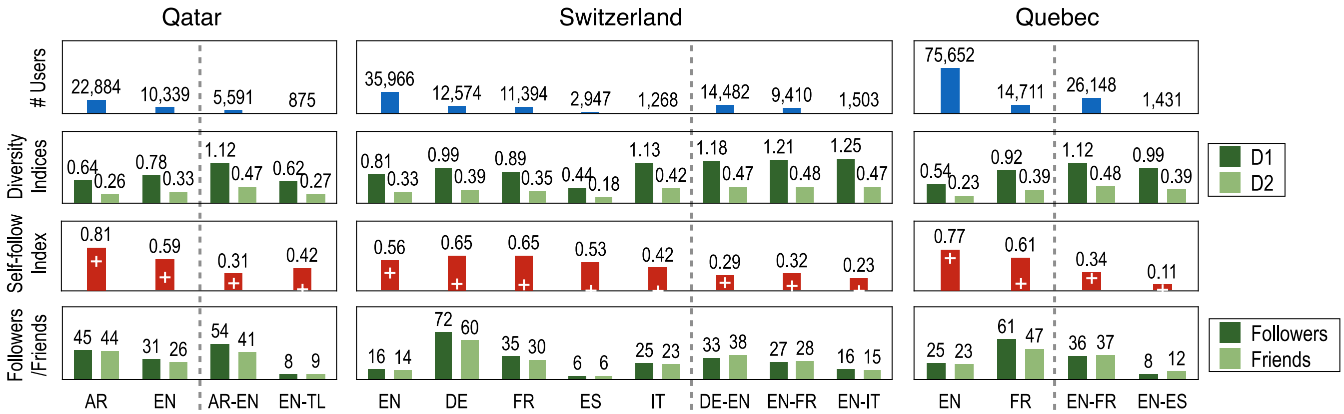

Figure 3: Network measurements for lingua groups in each multilingual region. Diversity indices D_1 (Eq. 2) and D_2 (Eq. 3) are calculated based on the distribution of the edges toward the lingua groups. *Self-follow Index* (Eq. 4) measures the probability that a new edge from a user would be directed to another user in the same lingua group. $+$ marks the baseline for *Self-follow Index*, based on purely random edges.

4. QUANTIFYING LANGUAGE USE AND NETWORK CHARACTERISTICS

In this section, we present a quantitative analysis for discovering social structure in the multilingual societies. After labeling each user by the language they tweet, we created a network for each region with nodes representing users and directed edges representing Twitter following relationships. Table 1 shows the statistics and network measurements for the five networks.

We first define a *lingua* as a mono- or multi-lingual combination of languages. We also define S is the set of all possible linguae, for instance, {EN, AR, AR-EN, DE-EN-FR, ...}. The "lingua group", U_i^r, is the set of users speaking lingua i in region r. We quantify the distribution of linguae for each region r using Shannon entropy:

$$H(r) = -\sum_{i \in S} p_i^r \log_2 p_i^r, \tag{1}$$

where p_i^r indicating the fraction of users speaking lingua i over the Twitter population of our data in region r. Higher $H(r)$ indicates that there is an even distribution of linguae over the population, and the entropy value for each region is shown in Table 1. As Figure 2 shows, Switzerland has the most diverse distribution of linguae, followed by Qatar and Quebec.

DATA STATISTICS			
Region	#Users	#Edges	#Tweets
Qatar	41,782	1,496,491	43,305,962
Switzerland	83,777	2,187,340	51,725,870
Quebec	110,434	3,074,463	73,676,397
NETWORK MEASUREMENTS			
Region	$H(r)^*$	Avg. degree	GCC** (%)
Qatar	1.403	71.63	99.50
Switzerland	2.195	52.22	98.77
Quebec	1.097	55.68	99.57

Table 1: Data Statistics and network measurements. Edges (directed) represent followings. See Eq. 1 for $^*H(r)$, for which higher value indicates that the distribution of languages in the region is less skewed. **GCC, node coverage of the greatest connected component over the network, is calculated with undirected edges.

Figure 1 shows the visualization of the networks. We used Gephi[3] software for plotting with the Yifan Hu graph layout algorithm. For the purpose of illustration, we highlighted the bilingual users. This visualization shows that monolingual users cluster together, while bilinguals are located between the two monolingual groups. To quantify this observation, we measured the diversity of outlinks for each lingua group. We first labeled and grouped users by the lingua. Then, we counted the edges between the groups. To calculate this, we define e_{ij}^r as the edge from user i to j in region r, pointing from Twitter follower i to the friend j. E_{ij}^r is the set of respective edges. Then we define P_{ij}^r as the proportion of the number of edges from i to j over the all outgoing edges from i. Formally,

$$|E_{ij}^r| = \sum_{a \in U_i^r, b \in U_j^r} e_{a,b}^r = \sum_{e \in E_{ij}^r} e \quad \text{and}$$
$$P_{ij}^r = |E_{ij}^r| / \sum_{k \in S} |E_{ik}^r|.$$

Diversity Measures. We first define D_1, based on Shannon Entropy of outgoing links for language subgroup i in region r, so that a lower D_1 indicates that outgoing edges will be more concentrated to a single lingua group. And we define D_2 using Simpson Index, the special version of inverse True diversity:

$$D_1(r,i) = -\sum_{j \in S} P_{ij}^r \log_2 P_{ij}^r \tag{2}$$

$$D_2(r,i) = 1 - \sum_{j \in S} (P_{ij}^r)^2 \tag{3}$$

D_2 is used in [21] as *Participation coefficient*, to measure how a node's connections are 'well-distributed', and from definitions, a lingua group with lower D_1 or D_2 has more concentrated outlinks into the small set of lingua groups. Since these metrics can only quantify the general shape of distribution, whether it is diverse or concentrated, we define *Self-follow index*, the probability of making a homogeneous connection when creating new edge from each user in a group. Higher intra-edges with lower inter-edges indicate the average user tends to follows another from the same language group. We call self-follow index of a lingua group as the average of self-follow index of all users in the group. The self-follow index

[3] http://gephi.org

of user a, region r and lingua group i is defined as

$$P(a) = (\sum_{b \in U_i^r} e_{a,b}^r) / (\sum_{j \in S} \sum_{b \in U_j^r} e_{a,b}^r). \quad (4)$$

Monolinguals Cluster Together. For all three regions, we found that monolingual groups consistently show lower D_1 scores than multilingual groups. As in D_1, monolingual groups have higher D_2 when compared to the multilingual groups in the same region except for the English monolinguals. Figure 3 shows both D_1 and D_2 for each group. We also found that monolingual lingua groups have higher self-follow index than any bilingual groups in all regions. The results from three diversity metrics and Intra-Inter edge ratio suggests that users in monolingual subgroups have a strong tendency to follow users inside of the same subgroup, while bilinguals do not.

Users of Local Language have Higher Influence. We explore the question of language use and social status, which we estimate simply with the number of followers, as studies have shown that tweets from a user with a high in-degree are more likely to be retweeted [11, 32]. We first look at the mean and median of the number of followers and friends of users in each lingua group within the network. This is to approximate the user's intra-region social status by excluding the effect toward the outside of the network in our data. To minimize the effects of outliers we removed the top and bottom 10% of users for the number of followers and friends. Figure 3 shows the average number of followers and friends for each lingua group in three regions, and median strictly followed the mean numbers. In all three regions and all lingua groups, we found that the number of friends is always larger than the number of followers. Also, for all three regions, users tweeting in the local language have more followers and friends, even when there are more English monolinguals in the dataset, such as in Switzerland and Quebec. This phenomenon shows that users tweeting in the local language exert higher influence within the regional network.

5. ANALYSIS OF BILINGUAL GROUPS

In this section, we analyze questions such as: Do bilinguals act as bridges between monolingual groups? Is there a pattern of language convergence where, say, when your audience contains a certain fraction of English-only speakers you switch to English?

Bilinguals and English Act as Bridges. The previous section showed that monolinguals form clusters. Now we analyze the mono- and bilingual bridges that glue multilingual societies together, as well as help to avoid language-ghettoization. We first visualize how monolingual and multilingual users follow each other. Figure 4 shows the ratio of follows among lingua groups in the three regions. A node represents a lingua group, and the size of a node corresponds to the relative number of users in that group. We only show nodes for the lingual groups that are represented in Figure 3. Only edges with weight higher 10% are shown to avoid visual clutter. To calculate the numbers underlying the figure, we first get the follow distribution toward lingua groups for each user, then averaged the distributions for all users in the same lingua group. We found that bilingual groups bridge monolingual groups. The key findings are (i) English acts as a hub language, meaning monolingual groups are connected through a X-EN bilingual group, or through the EN group, (ii) Bilingual group X-Y bridges two monolingual groups X and Y, (iii) In-group following takes the largest proportion for monolingual groups, and (iv) Monolingual users do not follow monolingual users of another language.

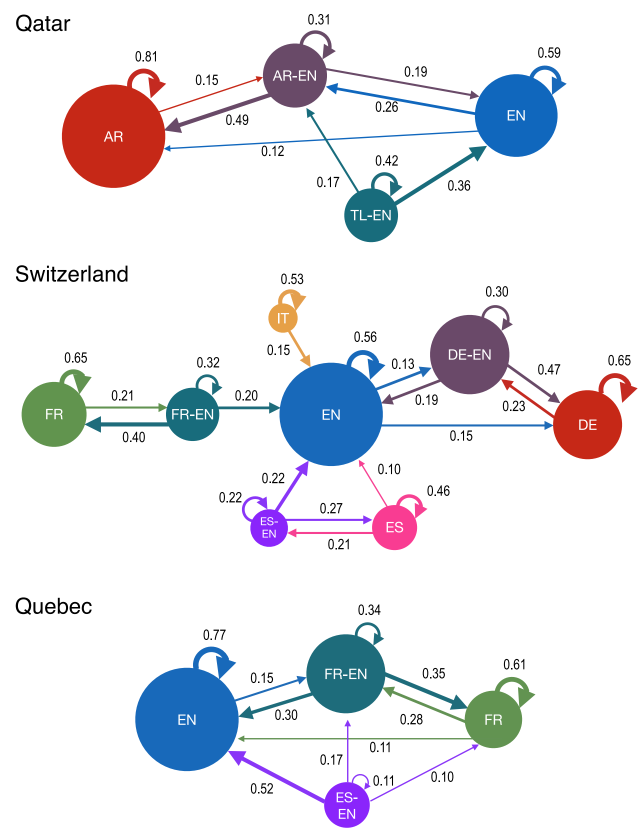

Figure 4: Following patterns among lingua groups. The color of an edge corresponds to the source color of the edge. Following distribution normalized for each user and averaged over group. The number for each edge corresponds to the ratio of the followings over all from source group. Edges having weight > 0.10 are shown.

For all three regions, we found that English users communicate with bilingual groups. Specifically, monolingual groups are strongly connected with respective bilingual groups, which in turn are connected to the EN group. Such connection property forms a star-shaped network with EN group as a hub. Our observation that English acts as a hub language revalidates the prior finding that English is used as a lingua franca in Twitter.

Language Convergence Consistent Across Regions. Given that the tweets are broadcast to every follower, how should multilinguals choose their language? A game-theoretic approach with an objective of "maximizing the audience" might predict that it requires the users to switch to the language of largest fraction. This could then quickly lead to a global convergence to a single lingua franca and pose a threat to the preservation of language. We investigate this issue by looking at the language distribution of bilingual users on Twitter. A user who at least occasionally tweets in different language has a choice and could use either language. How does their tweet mixing ratio, i.e., the fraction of tweets in English, depend on the mixing ratio of their followers, i.e., the average tweet mixing ratio of their followers, bilingual or not? If we were to observe a steep, threshold-like shape of the curve for English, where bilinguals predominantly use English as soon as a small fraction of their followers use only English, then this would spell trouble for the "native" languages.

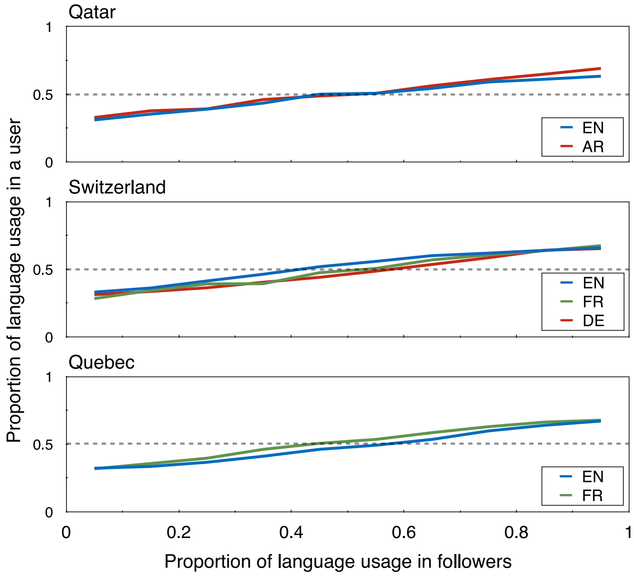

Figure 5: Trendline for the relation between language usage of a bilingual user and their monolingual friends. The X-axis indicates the average percentage of monolingual followers for a user and the language of interest. The Y-axis indicates the average percentage of a bilingual user's tweets in that language.

Figure 5 plots the language distribution of a bilingual user's monolingual friends on the x-axis and the language distribution of bilingual users with the corresponding language distribution among followers on the y-axis. To draw a trend line, we divide the x axis into ten bins, and average users' proportion of language usage for each bin. Users having < 5 followers in the induced network are filtered out. The observed pattern is very consistent among languages and geographical regions. Simply put, bilingual users "mimic" the language mix observed among their followers.[4] Though the equality does not hold for users at either end of the spectrum, say, with 90% English among their followers, this is likely an artifact as we only consider bilinguals. While it is still possible that a large number of users have already been converted to monolingual users, The very smooth and consistent pattern suggests that the language conversion process is more gradual than one might have expected.

Bilinguals Post Different Stories in Different Language. To gain a deeper understanding of the role of bilingual users in a multilingual society, we analyze the contents of the tweets from bilingual users. To observe any systematic differences in language use, we use a parallel set of tweets containing the same hashtags in two languages, train a topic model on those tweets to reveal the differences in the topics of the same hashtags. For this analysis, we train latent Dirichlet allocation model [7] to discover the topics and analyze any differences in information depending on the language. We did not use the Polylingual topic model[27], as it requires a corpus of documents in different languages with similar topics. After running LDA for each language and we translated the top words into English. We avoid the topic alignment problem by using the set of translated hashtag pairs. For instance, we use (#suisse - #switzerland) for EN-FR in Switzerland, and respective country hashtag pairs for other regions. We set the number of topics for each set, $k = 10$, and after fitting the model we use Google translate service to translate words into English. We set $\alpha = 50/k$ and $\beta = 0.01$ for LDA. Table 2 shows the part of topics from two regions. We found that from all bilingual groups in three regions, bilingual users post *informational and political* tweets for the local audience in local language. They, on the other hand, post *events, tourism, photography*, and other *leisure-related* tweets in English for the non-local audience. These results show that our methodology of identifying bilingual Twitter users and analyzing the topics of their tweets can reveal the semantics of communications among multiple language speakers in a multilingual society.

[4]We observed nearly identical plots when we also included bilingual followers for the x-axis and "bucketed" them according to their tweet mixing ratio.

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

Table 2: Part of the topics discovered from tweets containing country hashtags and are posted by bilinguals. Topics are manually labeled from the top words. We translated the most frequent words into English. We do not display stopwords and region names. For all three regions, tweets containing local language hashtag are mainly of informative/political/debatable topics, while tweets containing English hashtag are event/tour/enjoyment topics.

6. CONCLUSION

We presented a large-scale computational analysis of language use and network characteristics of the language-based groups in multilingual societies using Twitter data. Using the extensive set of tweets from monolingual and bilingual users from Qatar, Switzerland, and Quebec, we first discovered that monolingual users cluster together, while bilinguals do not. Then, we revealed that users speaking local language have more influence than others. Additionally, we have shown that, surprisingly, the language-mixing ratio of bilingual users closely mirrors the mix of their followership. Then we showed that bilinguals bridge between monolinguals with English as a hub, while monolinguals tend not to directly follow each other. Finally, with the statistical topic model, we discovered that bilinguals express informative/political/debatable topics in a local language, while posting event/tour/enjoyment topics on the English.

Acknowledgements

This research was funded by the MSIP (Ministry of Science, ICT & Future Planning), Korea in the ICT R&D Program 2014.

7. REFERENCES

[1] D. M. Abrams and S. H. Strogatz. Linguistics: Modelling the dynamics of language death. *Nature*, 2003.

[2] J. Androutsopoulos. Language choice and code-switching in german-based diasporic web forums. *The multilingual Internet*, 2007.

[3] J. Anis. Neography: Unconventional spelling in french sms text messages. *The multilingual Internet*, 2007.

[4] A.-S. Axelsson, Å. Abelin, and R. Schroeder. Anyone speak swedish? tolerance for language shifting in graphical multi-user virtual environmnets. *The multilingual Internet*, 2007.

[5] G. Bailey, J. Goggins, and T. Ingham. `http://mlm.humanities.manchester.ac.uk/reports/Twitter%20and%20Language%20Diversity.pdf`.

[6] S. Bergsma, P. McNamee, M. Bagdouri, C. Fink, and T. Wilson. Language identification for creating language-specific twitter collections. In *LASM*. ACL, 2012.

[7] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *JMLR*, 2003.

[8] P. Bourdieu. *Distinction: A social critique of the judgement of taste*. Harvard University Press, 1984.

[9] S. Carter, W. Weerkamp, and M. Tsagkias. Microblog language identification: Overcoming the limitations of short, unedited and idiomatic text. *Language Resources and Evaluation*, 2013.

[10] X. Castelló, V. M. Eguíluz, and M. San Miguel. Ordering dynamics with two non-excluding options: bilingualism in language competition. *New Journal of Physics*, 2006.

[11] M. Cha, H. Haddadi, F. Benevenuto, and P. K. Gummadi. Measuring user influence in twitter: The million follower fallacy. *ICWSM*, 2010.

[12] S. Climent, J. Moré, A. Oliver, M. Salvatierra, I. Sànchez, and M. Taulé. Can machine translation enhance the status of catalan versus spanish in online academic forums? *B. Danet & S. Herring (Eds.), The multilingual Internet*, 2007.

[13] X. Daming, W. Xiaomei, and L. Wei. 15 social network analysis. *The Blackwell guide to research methods in bilingualism and multilingualism*, 2008.

[14] J.-M. Dewaele. Blistering barnacles! what language do multilinguals swear in? *Estudios de Sociolinguistica*, 2004.

[15] N. C. Dorian. *Language death: The life cycle of a Scottish Gaelic dialect*. University of Pennsylvania Press Philadelphia, 1981.

[16] N. C. Dorian. *Investigating obsolescence: Studies in language contraction and death*. Cambridge University Press, 1992.

[17] M. Durham. Language choice on a swiss mailing list. *JCMC*, 2003.

[18] J. A. Fishman. *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages*. Multilingual matters, 1991.

[19] S. Gal. *Language shift: Social determinants of linguistic change in bilingual Austria*. Academic Press New York, 1979.

[20] L. A. Grenoble and L. J. Whaley. *Endangered languages: Language loss and community response*. Cambridge University Press, 1998.

[21] R. Guimera and L. A. N. Amaral. Functional cartography of complex metabolic networks. *Nature*, 2005.

[22] C. K.-M. Lee. Linguistic features of email and icq instant messaging in hong kong. *The multilingual Internet*, 2007.

[23] W. Li. *Three generations, two languages, one family: Language choice and language shift in a Chinese community in Britain*. Multilingual Matters, 1994.

[24] M. Mainguy, Y. Nakai, and M. Takayama. `http://mlm.humanities.manchester.ac.uk/reports/Mapping%20Mancunian%20Multilingualism%20on%20Twitter.pdf`.

[25] E. Manley. `http://urbanmovements.co.uk/2012/10/23/detecting-languages-in-londonstwittersphere/`.

[26] L. Milroy. *Language and social networks*. B. Blackwell, 1987.

[27] D. Mimno, H. M. Wallach, J. Naradowsky, D. A. Smith, and A. McCallum. Polylingual topic models. In *EMNLP*, pages 880–889, 2009.

[28] J. W. Minett and W. S. Wang. Modelling endangered languages: The effects of bilingualism and social structure. *Lingua*, 2008.

[29] S. S. Mufwene. *Language evolution: Contact, competition and change*. Continuum International Publishing Group, 2008.

[30] T. Poell and K. Darmoni. Twitter as a multilingual space: The articulation of the tunisian revolution through #sidibouzid. *NECSUS*, 2012.

[31] M. S. Schmid. *Language attrition*. Cambridge University Press Cambridge,, UK, 2011.

[32] B. Suh, L. Hong, P. Pirolli, and E. H. Chi. Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In *SocialCom*. IEEE, 2010.

[33] R. Wardhaugh. *Languages in competition: Dominance, diversity, and decline*. B. Blackwell, 1987.

[34] M. Warschauer, G. R. E. Said, and A. G. Zohry. Language choice online: Globalization and identity in egypt. *JCMC*, 2002.