DOI: 10.30050/CLLT.202412_4(2).0001

Making Meanings at Meetups: Gender Classification with Transcribed Bilingual Texts

Chun-Yi Peng
Department of Modern Languages
Borough of Manhattan Community College
City University of New York

Abstract

New York City has a dynamic meetup culture where people gather for shared interests. Some of the most popular meetups are focused on language learning. Learners from diverse backgrounds and proficiency levels come together regularly to practice the target language(s). Due to the diversity of meetup participants and the multitude of languages spoken in New York City, these language meetups provide fertile research ground for linguistic research.

However, there has been limited research regarding how gender plays a role in SLA within the context of meetup scenes. Sociolinguists have long studied gender differences in language use across various social contexts, but it remains unclear whether these findings or hypotheses are applicable to second language acquisition. Drawing upon transcribed bilingual data, this study aims to 1) explore gender differences in second language use, including word frequency, code-switching, ratio of target language use, sentiment, and emotions, and 2) build a computational model with these hypotheses as features to predict the gender of speakers.

Keywords: meetups, gender, second language acquisition, machine learning

1. Introduction

New York City boasts a dynamic meetup culture where people gather for shared interests. One of the most popular meetups is centered around language learning. Learners from diverse backgrounds and proficiency levels regularly convene to practice their target language(s). Given the richness of participants and the multitude of languages spoken in New York City, these language meetups serve as fertile ground for research on second language acquisition (SLA), communication studies, and discourse analysis.

Despite the vibrant meetup scene, there has been limited exploration, especially in the context of SLA among adult language learners in large urban environments. Thus, this study collected L2 Chinese speech data and transcribed it into texts to scrutinize gendered language use within a Chinese meetup group. The objectives of this study are twofold: 1) to delve into gender differences in second language use, encompassing aspects such as word frequency, code-switching, ratio of target language use, sentiment and emotions, and 2) to construct a computational model utilizing these hypotheses as features to predict the gender of speakers.

2. Literature review

The ongoing discourse surrounding language and gender has spurred linguistic investigations from diverse perspectives since Labov's pioneering studies (1990, 2001). His *principles of linguistic change* lay the foundation for subsequent research on language and gender (Labov, 2001):

- Principle I: In stable variation, women use more of the standardized variant than men do.
- Principle Ia: In changes from above, women favor the incoming prestige variant more than men.
- Principle II: In changes from below, women are most often the innovators.

Sociolinguists later delve into the examination of social practices and linguistic features associated with manliness or masculinity, as evident in the works of scholars such as Eckert & McConnell-Ginet (1992, 1999), Lawson (2011, 2013, 2020), and Kiesling (2018). Discourse analysts, on the other hand, scrutinize how masculinity is constructed through discourse or media representation in various social contexts, as demonstrated by the studies of Seidler (2006), Ehrlich & Levesque (2011), Milani (2013), Baker & Levon (2016), and Peng & Garcia (2020). Moreover, computational linguists contribute to the discussion by analyzing textual data from social media (Bamman et al., 2014; Miller, 2021; Sap et al., 2014; Voigt et al., 2018) and literary works (Argamon et al., 2003) to predict the gender of the author or to examine gendered language use.

In the realm of computational linguistics, Bamman et al. (2014) utilized a corpus of 14,000 Twitter users to discern the gender of authors. They identified key gender-identifying features, such as pronouns and emotion words for females, and numbers and technology words for males. Notably, their findings revealed that misclassified authors were often connected through an online social network where individuals were more likely to mention or tag others with the same gender. Sap et al. (2014) took a different approach by deriving predictive lexica (words and weights) for age and gender through regression and classification models based on word usage in Facebook, blogs, and Twitter data with associated demographic labels. They found that the lexica maintained reasonable accuracy with smaller datasets, such as 20 messages, while suggesting that other approaches might be more suitable for larger datasets.

Voigt et al. (2018) delved into communication directed towards individuals of different genders, compiling a multi-genre corpus of over 25 million comments from five socially and topically diverse sources, all tagged for the gender of the addressee. They employed a simple unigram logistic regression model to predict the gender of the source. Results indicated that responses to women were often biased and more likely to be emotive. Research on the roles of gender in second language studies has followed the general trend of sociolinguistic studies. Over the past two decades, it has undergone a paradigm shift from the focus on binary differences to more and more diverse concept of gender identity and practices (Cameron, 2005; Menard-Warwick, 2012; Pavlenko et al., 2001).

Methodologically, however, little has been done in second language acquisition to explore the role of gender computationally. Thus, this study represents a preliminary attempt to utilize computational models to examine the role of gender in Second Language Acquisition (SLA). This innovative perspective contributes to the broader understanding of the intersection between gender and linguistic phenomena, shedding light on an area that has remained relatively unexplored in the realm of second language acquisition.

3. Data and participants

I collected data[1] by participating in a Chinese-English exchange meetup in Midtown Manhattan once a week on Thursday nights. Verbal consents were received from the participants to record the meetup. The language exchange meetup had an informal structure, without apparent leaders or rules. Therefore, it was an ideal site for collecting spontaneous speech data and natural social interaction. The number of participants varied each week, but there were usually 15-20 people each week at the meetup. The data was collected from 32 male and 8 female participants (N = 40) over 5 meetup sessions before the pandemic between April to June 2019. The corpus for this study consists of transcribed audio data and written notes by the researcher. A total of 1403 sentences were extracted for analysis. All sentences were in the format of 'name: quote'.

Participants

Most participants were in their 20s. They came from various linguistic backgrounds, with speakers of English, Polish, German, Cantonese, etc. Speakers of intermediate or advanced proficiency level often began speaking in their non-native language, but after a few hours often relapsed to their most comfortable language for long stretches, often clustering with others who spoke the same language as them and practicing the foreign language together. Low-proficiency speakers tended to talk mainly in their own languages, or quietly listen on the periphery, sometimes jumping in or reacting to funny information.

[1] All the data and codes for this project can be accessed on GitHub.

High-proficiency learners sometimes act as teachers (of their native language) or conversation partners for others trying to learn that language.

When conversations were primarily among a pair or group of non-native mandarin speakers, they joked or showed off their knowledge about the funny little cultural, linguistic, or societal fact they picked up about China - such as idioms, useful apps, or surprising facts about ‘the way things are’ there. On the contrary, when talking to a Chinese person, the Chinese learners often ask more questions to learn from their perspective. If the Chinese attendee is very new to the US, the conversation may turn more towards helpful practical tips about how to get on in New York City. Learners at the beginning level tended to be most drawn to these conversations, which gave them a chance to code-switch to English while still sharing their valuable knowledge.

4. Methods

I first explored the data by examining the descriptive statistics to get a sense of the gender differences in the data, including 1) frequency of words, 2) frequency of code-switching, 3) target language use, 4) the sentiment polarity scores and 5) use of emotion words. Then, I used these as features to build computational classifiers that extract features from utterances to predict the gender of speakers. Finally, I evaluated these classifiers to identify the best model.

Preprocessing

The field note consisted of the researcher’s notes and transcriptions of conversations at the meetups. The conversation data was formatted as ‘name: direct quotes’. During preprocessing, 1,403 direct quotes were extracted from the corpus using regular expressions. Each direct quote was labeled with binary gender (i.e., male or female) using the name information. Before inputting into a classifier, the data was formatted as ‘name: direct quotes’ (i.e., one column for names and one column for quotes). For exploratory data analysis, I calculated word frequencies, including their overall frequencies and frequencies by gender. Punctuation and “stop words” were removed from the word count. The results were visualized in figures 1 and 2 in the next section. Additionally, I calculated the frequencies of code-switching by gender, followed by a one-sample T-test. Sentiment analysis was also conducted to assess how positively male and female speakers expressed themselves, using the VADER sentiment analysis tools (Hutto & Gilbert, 2014).

Training the classifiers

After exploring the data, I built a "feature extraction" function to tokenize the "direct quotes" and extract features. The feature extraction function first tokenized both English and Chinese words, and then extracted five features: gender-distinguishing words, code-switching, target language use, sentiment polarity score, and emotion scores. These features are defined as follows:

1) Word frequency = the total number of occurrences of a word in the corpus
2) Frequency of code-switching = times of language change in a sentence
3) Target language[2] use = number of tokens in the target language/ total number of tokens
4) Sentiment polarity scores[3] = the "compound" score calculated by VADER
5) Emotional score = "raw emotion score" calculated by NRCLex[4]

The extracted features of each sentence were vectorized. The vectorized data was then split into training, developing, and testing sets by the ratio of 0.8: 0.2: 0.2. The following table summarizes the basic statistics of the data:

Table 1 Data size by gender

Label	**Size**
Male sentences	1186
Female sentences	217
Total	1403

[2] The target language in this study is Mandarin.

[3] The compound score in VADER sentiment analysis tools is the sum of the positive, negative, and neutral scores for each word in the lexicon. A compound score >= .5 shows a positive sentiment; A score between -.5 and .5 shows a neutral sentiment; a score < .5 shows a negative sentiment.

[4] https://pypi.org/project/NRCLex/

Table 2 Vectorized data

Data type	**Size**
Training	6732
Developing	846
Testing	840
Total	8418

Four classifiers were constructed using these vectorized features to predict the gender of each utterance, encompassing a baseline classifier, a naïve Bayes classifier, a logistic regression classifier, and a neural network classifier. The initial three classifiers were developed using Scikit Learn (Pedregosa et al., 2011), while the neural network classifier was implemented with PyTorch. All classifiers underwent training on the designated training dataset.

The baseline model was established using Scikit Learn’s DummyClassifier, which generates predictions without considering input features, serving as a baseline against other classifiers. Often used for text data, the naïve Bayes classifier is a probabilistic classifier that assumes features are independent. Logistic regression is another model commonly used for binary classification tasks. In this study, the logistic regression model was seeded to ensure reproducibility, and hyperparameters were tuned. The parameters ‘C’ and ‘penalty’ were adjusted on the development set, exploring possible values such as C $\in$ [.001, .01, .1, .2, .5, 1., 2., 5., 10., 20., 50., 100] (‘C’: .001, ‘penalty’: ‘l1’)[5].

The neural network classifier adopts a feedforward neural network structure, comprising three layers. The first two layers consist of a hidden function (linear), an activation function (ReLU), and a dropout function. The final layer includes an output function (Sigmoid). Ultimately, the classifiers were assessed using the testing set (see figure 1).

[5] In logistic regression, the term ‘C’ refers to the inverse of regularization strength, and ‘penalty’ refers to the regularization term added to the loss function to prevent overfitting.

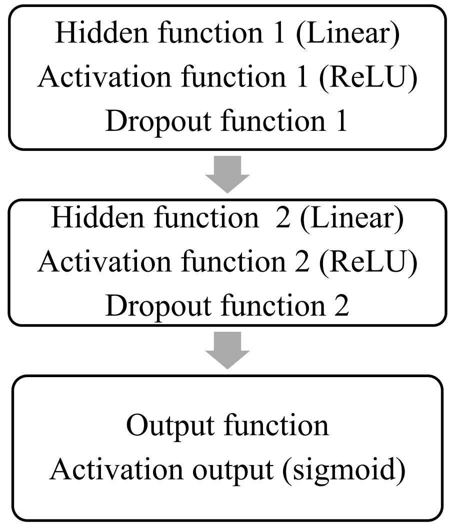

Figure 1 Neural Network Classifier

5. Results and evaluation

5.1. Word frequency

The most frequent words for male and female participants are listed in Table 1. The results are shown in Table 1, also visualized in Figures 1 and 2.

Table 3 Words with highest frequency

Rank	**Words by males**	**Counts**	**Words by females**	**Counts**
1	I	427	I	43
2	Yeah	162	Yeah	29
3	oh	109	的	23
4	like	101	我	18
5	Uh	84	know	13
6	的	72	Uh	13
7	know	71	你	12
8	people	63	Huh	11
9	Chinese	61	是	11
10	你	54	China	11

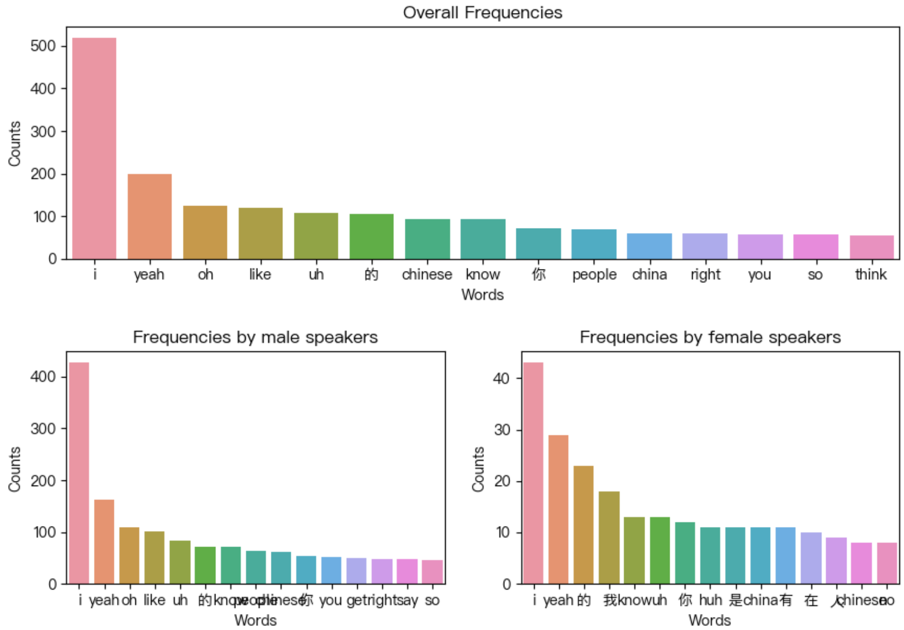

Figure 2 Word frequencies

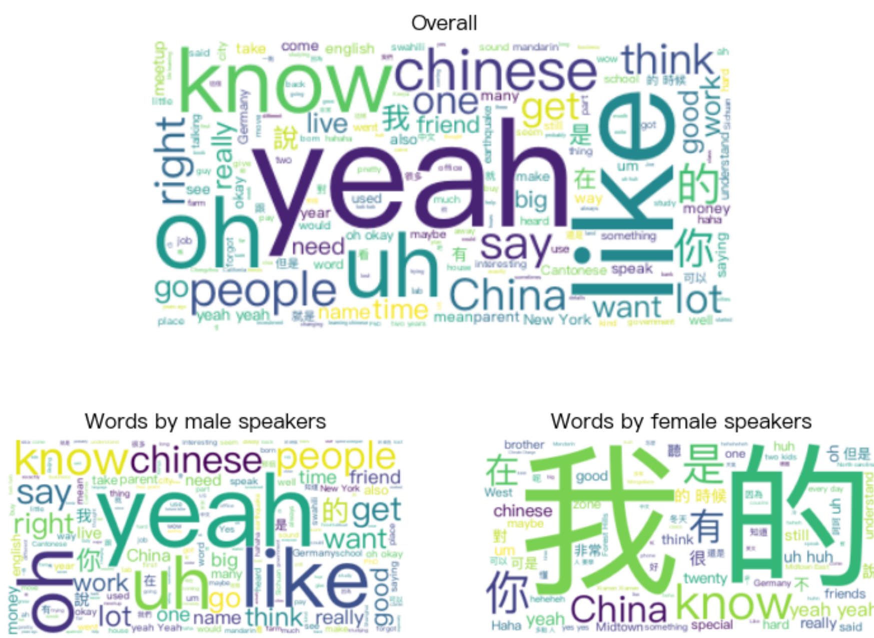

Figure 3 Word clouds

Based on the results of word frequency count, the following words are chosen as gender distinguishing words and are listed as the first feature in the model:

- Male words: “people”, “get”, “Chinese”, “like”, “get”, “right”, “oh”
- Female words: “我”, “是”, “有”, “在”, “人”, “China”

Words frequently used by male speakers are all in English whereas those by females are mostly in Chinese, suggesting that females make more attempts to use the target language. Below are examples of these gender-distinguishing words in sentences:

Sentences by males

> “All I know is that it was hilly and there were lots of Chinese people. But In LA there are way more Chinese people overall than in New Jersey so I bet anywhere I went would seem like tons of Asians...”
> “Yeah, but in Cupertino if I want to get to my friend’s house, you know … I gotta take a bus there … take a couple buses in a couple hours you can make it.”

Sentences by females

> 我的名字是...
> (My name is…)
> 德國不錯，可是天氣不好，有點冷。冬天很冷。
> (Germany is not bad, but the weather is not good. It’s a little cold. The winter is very cold.)

5.2. Frequency of code-switching

As for code-switching, male speakers code-switched 366 times, or 31 times per 100 sentences, while female speakers code-switched 42 times, or 19 times per 100 sentences. However, the difference is not statistically significant ($p > .05$). I will test whether this feature increases or decreases the performance of the model. The results are summarized in Table 4.

5.3. Ratio of target language use

As shown in Table 4, male participants use the target language (8.52%) significantly less than the female participants (25.58%) ($p < .05$). Therefore, the ratio of target language use is included as a feature in the model.

5.4. Sentiment analysis

Sentiment analysis was conducted using Vader-Sentiment-Analysis (Hutto & Gilbert 2014). I calculated the “compound score”[6] of each sentence. The average compound score for sentences by male participants was .11, female participant .08 (p > .05). A compound score of .5 or higher indicates a positive sentiment, while a score lower than -.5 indicate a negative sentiment. Anything between -.5 and .5 reflects a neutral sentiment. Since the difference is not statistically significant, I will also test whether this feature increases or decreases the performance of the modal.

5.5. Emotion words

Although Bamman et al. (2014) suggests that emotion terms (sad, love, glad, etc.) and emoticons are commonly associated with females, the data show that sentences by male participants received higher emotion score than those by their female counterparts. Below are examples of sentences which receive the highest emotion scores:

Example 1 (in Mandarin by a male speaker, translation by the author. Emotion score: 22)

> 在我住的鎮，大約有五分之四的人死亡。他們在山裡有自己的家。我們在那裡工作，但山塌了。一樓塌了。他們離開了一樓 - 但我們的樓層保持不變。一樓徹底被壓垮了。我們附近的另一棟辦公大樓被壓垮了。但它是空的。

[6] https://vadersentiment.readthedocs.io/en/latest/pages/about_the_scoring.html

(In my town there were about four of five people dead. They had their own home inside the mountains. We worked there, but the hill crushed. The first floor crushed. They left the first floor - but our floor stays the same. The first floor was completely crushed. Another office building near us got crushed. but it was empty.)

Example 2 (in English by a male speaker. Emotion score: 19)

"No, I just walked in one day and was like 'do you need a dishwasher?' and then I started serving drinks too. It's good money too, I mean, um… I don't know, there's a lot of locals in the winter time and since it is so slow you don't need anyone else to work so you can do all the cooking and the serving and there's no one else working but you so you basically get all the day's pay for all the jobs."

Example 3 (in Mandarin by a female speaker, translated by author. Emotion score: 10)

因為，他們沒有一胎化政策。他們沒有，他們沒有。
(Because, they don't have one child policy. They don't have it, don't have it.)

Table 4 Summarizes the statistics of the features for male and females

	Male	Female	*p*-value
Times of code-switching (per 100 sentences)	30.68	19.35	0.0679
Target language use (%)	8.25	25.58	<.001***
Polarity score	.11	.08	0.14
Emotion score	1.46	.74	<.001***

5.6. Evaluation

After data exploration, four classifiers were built on the training data: a dummy classifier (baseline), a logistic regression classifier, a Naïve Bayes classifier, and a neural network classifier. All models were evaluated on the testing data set. Due to the imbalanced nature of the dataset (i.e. mostly utterances by males), models were evaluated by both precision and accuracy scores as summarized in Table 5. The baseline, logistic regression, and neural network classifiers show the same precision and accuracy scores (.8714). The Naïve Bayes classifier has the highest precision score, but the lowest accuracy score compared the other three models. Neural network does not outperform other models, possibly due to the small size of the dataset.

Table 5　Model comparison

Model	**Precision**	**Accuracy**
Baseline	0.8714	0.8714
Logistic regression	0.8714	0.8714
Naïve Bayes	0.8897	0.8571
Neural network	0.8714	0.8714

Most misclassified sentences are labeled as female but are mistakenly classified as male. This is particularly evident with the dummy classifier, which tends to classify almost everything as male due to a disproportionate amount of data coming from male participants. However, when compared to other classifiers, the Naïve Bayes classifier tends to misclassify more male sentences as female. Here are some misclassified sentences from the Naïve Bayes classifier where male sentences were labeled as female:

Table 6　Misclassified sentences from the Naïve Bayes classifier

Label	**Prediction**	**Sentence**
male	female	Oh yeah…
male	female	Well, I think there are just a lot of Fujianese people in New York in general. So, like in Flushing and stuff… yeah. Yeah so… it’s kinda an um...

Sentences with only one or two words are more likely to be misclassified as shorter sentences provide less information for model classification. Here are some misclassified examples from the Logistic regression classifier:

Table 7 Misclassified sentences from the Logistic Regression classifier

Label	**Prediction**	**Sentence**
female	male	really?!
female	male	PhD, yes.
female	male	Exactly.

Although the goal of this study is to propose a computational method to analyze second language data, several caveats must be considered regarding the limitations of this approach. First, despite the increasing recognition of gender as a spectrum rather than a binary category (Cameron, 2005), this project adopts a binary approach to gender for ease of classification by computational models. Given this approach, the gender ratio of participants in this study is imbalanced, with a disproportionately higher number of male participants (n = 32) compared to female participants (n = 8) in the corpus. Thus, the results should be interpreted with caution due to the imbalanced gender ratio among the participants. Second, this project does not control for participants' proficiency levels and native languages. Second language speakers may employ different communication strategies at various phases of language learning. For adult learners, their native languages may also play a role in their second language use. However, this study does not control for either participants' proficiency levels or native languages due to the relatively small size of the dataset. Since the data for this study comes from spontaneous speech by participants of New York City meetups, it is difficult, if not impossible, to measure each participant's proficiency level. Many of the participants also speak multiple languages, which also makes it challenging to examine the role of native languages in second language use.

6. Conclusion

This study is a preliminary attempt to construct gender classifiers using transcribed bilingual text data. It draws upon 1,403 sentences from 32 male and 8 female meetup participants to build four gender classifiers: a dummy classifier, a logistical regression, a naïve base, and a neural network classifier. Strictly speaking, none of the selected features significantly improve the precision or accuracy of the models beyond that of the baseline model. The results of this project could have benefitted from a much larger dataset. With a small dataset, the neural network model also fails to improve the precision or accuracy of the classifier. Another possibility could be that the selected features are not gender discriminative. Among the five selected features, only ‘use of key words’ and ‘target language use’ are significant predictors of gender. ‘Frequency of code-switching’ and ‘polarity score’ have a weaker effect on the precision and accuracy of the models.

Another drawback is related to the lack of context in the analyses. Since discourse analysis unpacks the texts’ cultural meaning and power relationships, text analysis should be a human undertaking (van Dijk, 1980). The bag-of-words technique treats words as independent of the original word order and dislocates content from its context. It is helpful for document classification, frequency counts, or word clustering, as it reduces complexity. However, if the context in which the words are used is an important component of the text, removing words from this context eliminates crucial aspects of the data (Sinclair, 1991; Schuelke-Leech & Barry, 2017).

This study merits future work in two directions. Methodologically, in addition to a much larger dataset, the neural network classifier could be improved by adding more layers to the neural network model or by adopting a Long Short-Term Memory neural network model. It would also be interesting to explore the data with unsupervised models, such as Clustering or Latent Dirichlet Allocation. Instead of testing preconceived assumptions, unsupervised models can identify certain patterns or themes from the text data. Linguistically, incorporating discourse nuances into the models may also enhance the performance of the classifiers. For example, under the framework of Conversation Analysis (Sacks et al., 1974; Schegloff, 2007; Heritage, 2015), there are communication strategies that might also be gender-marking, such as

1. *Turn taking*: Who is taking the floor in the conversation? Under what circumstances does turn taking happens?
2. *Sequence organization*: When do silence and pauses happen? How do people transition between topics?
3. *Repair*: When do speakers self-correct? When do speakers repeat what they said?

Traditionally, these communication strategies are studied with data from native speakers. However, research has found that these communication strategies are also helpful for second language learners to emulate native-like authentic speech. Therefore, analyzing these strategies might shed light on the natural social interaction in an SLA context. Through the computational analyses of text features, this work encourages more future research on how social variables like gender is constructed and performed in the process of SLA.

Acknowledgements

Support for this project was provided by a PSC CUNY Award (#64389-00 52), jointly funded by The Professional Staff Congress and The City University of New York. I am also deeply grateful to Dr. Nicholas Garcia for his assistance in data collection and transcription. All errors are my own.

References

Argamon, S., Koppel, M., Fine, J., & Shimoni, A. R. (2003). Gender, genre, and writing style in formal written texts. *Text & Talk, 23*(3), 321–346. https://doi.org/10.1515/text.2003.014

Baker, P., & Levon, E. (2016). 'That's what I call a man': Representations of racialized and classed masculinities in the UK print media. *Gender & Language, 10*(1), 106–139.

Bamman, D., Eisenstein, J., & Schnoebelen, T. (2014). Gender identity and lexical variation in social media. *Journal of Sociolinguistics, 18*(2), 135–160. https://doi.org/10.1111/josl.12080

Cameron, D. (2005). Language, gender, and sexuality: Current issues and new directions. *Applied Linguistics*, *26*(4), 482–502. https://doi.org/10.1093/applin/ami027

Eckert, P., & McConnell-Ginet, S. (1992). Think practically and look locally: Language and gender as community-based practice. *Annual Review of Anthropology, 21*, 461–490.

Eckert, P., & McConnell-Ginet, S. (1999). New generalizations and explanations in language and gender research. *Language in Society, 28*(2), 185–201.

Ehrlich, S., & Levesque, S. (2011). The strategic marginalization of working class masculinity in a batterers' treatment programme. *Gender & Language, 5*(2), 271–301.

Heritage, J. (2015). Well-prefaced turns in English conversation: A conversation analytic perspective. *Journal of Pragmatics, 88*, 88–104.

Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media, 8*(1), 216–225.

Kiesling, S. F. (2018). Masculine stances and the linguistics of affect: On masculine ease. *NORMA 13*(3–4), 191–212.

Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change, 2*(2), 205–254. doi:10.1017/S0954394500000338

Labov, W. (2001). *Principles of linguistic change, volume 2: Social factors.* New Jersey: Wiley-Blackwell.

Lawson, R. (2011). Patterns of linguistic variation among Glaswegian adolescent males. *Journal of Sociolinguistics, 15*(2), 226–255.

Lawson, R. (2013). The construction of 'tough' masculinity: Negotiation, alignment and rejection. *Gender and Language, 7*(3), 369–395.

Lawson, R. (2020). Language and masculinities: History, development, and future. *Annual Review of Linguistics, 6*, 409–434.

Menard-Warwick, J. (2012). Gender and second language acquisition. In *The encyclopedia of applied linguistics*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781405198431.wbeal0447

Milani, T. M. (2013). Are 'queers' really 'queer'? Language, identity and same-sex desire in a South African online community. *Discourse & Society, 24*(5), 615–633.

Miller, S. (2021). *When misclassification is misgendering: Gender prediction in the context of trans identities* (Master's thesis). CUNY Graduate Center. https://academicworks.cuny.edu/gc_etds/4203

Pavlenko, A., Blackledge, A., Piller, I., & Teutsch-Dwyer, M. (2001). *Multilingualism, second language learning, and gender*. Berlin, New York: De Gruyter Mouton.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research,12*, 2825–2830.

Peng, C. Y., & Garcia, N. (2020). Mediatized Taiwanese Mandarin: A text-mining approach to speaker stereotypes. *Open Linguistics, 6*(1), 611-625. https://doi.org/10.1515/opli-2020-0035

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*(4), 696–735.

Sap, M., Park, G., Eichstaedt, J., Kern, M., Stillwell, D., Kosinski, M., Ungar, L., & Schwartz, H. A. (2014). Developing age and gender predictive lexica over social media. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1146–1151. https://doi.org/10.3115/v1/D14-1121

Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge: Cambridge University Press.

Schuelke-Leech, B., & Betsy B. (2017). Philosophical and methodological foundations of text data analytics. In M. Dehmer, & F. Emmert-Streib (Eds.), *Frontiers in data science* (pp.147–169). London: Taylor & Francis Group.

Seidler, V. (2006). *Transforming masculinities: Men, cultures, bodies, power, sex and love.* London: Routledge.

Sinclair, J. (1991). *Corpus, concordance, collocation.* Oxford: Oxford University Press

van Dijk, T. A. (1980). *Macrostructures: An interdisciplinary study of global structures in discourse, interaction, and cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates. https://www.routledge.com/Macrostructures-An-Interdisciplinary-Study-of-Global-Structures-in-Discourse/Dijk/p/book/9780367112554

Voigt, R., Jurgens, D., Prabhakaran, V., Jurafsky, D., & Tsvetkov, Y. (2018). RtGender: A corpus for studying differential responses to gender. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. LREC 2018, Miyazaki, Japan. https://aclanthology.org/L18-1445

〔收稿：2023.07.17；修訂：2024.01.01；接受：2024.07.24〕

彭駿逸 Chun-Yi Peng
紐約市立大學曼哈頓區社區學院當代語言系
Modern Languages Department
Borough of Manhattan Community College
City University of New York
cpeng@bmcc.cuny.edu

聚會脈絡中的意義建構：以轉錄的雙語文本進行性別分類

彭駿逸

當代語言系

紐約市立大學曼哈頓區社區學院

摘要

紐約市有一個特別的聚會（meetup）文化，是由一群完全不認識的人，透過網路的連結，參加自己感興趣的聚會。其中，語言學習聚會是最受歡迎的聚會主題之一。來自不同文化背景和語言程度的學習者，定期聚在一起練習目標語言。由於紐約市聚會參與者的文化及語言多樣性，這些語言學習聚會的語料不僅為民族誌（ethnography）研究提供了豐富的研究基礎，還可用於第二語言習得、口語傳播和話語分析相關的研究。

然而，目前還沒有太多的研究聚焦於聚會脈絡下，性別在第二語言習得中所扮演的角色。社會語言學家長期以來一直研究性別在不同社會脈絡下，對於語言使用所造成的差異，而這些發現或假設是否適用於第二語言習得卻較少受到關注。本研究利用在中文／英文語言聚會所搜集的雙語資料：1）探討第二語言使用中的性別差異，包括詞頻、語碼轉換、目標語使用比例和情感等；2）建立一個計算模型，將這些假設作為特徵來預測說話者的性別。

關鍵字：聚會、性別、第二語言習得、機器學習