Talk like Mom: Sociolinguistic Interpretation of a Representation of Social Relationship

Pakawat Nakwijit

Introduction

Language is not only a medium to exchange information but also contributes to the performance of action within the interaction. Language serves simultaneously as a reflection of the relative positioning of speakers to their conversation partners as well as actions that accompany those positions (Gilbert & Karahalios, 2009). Many linguistic features have been acknowledged to correlate well with the closeness of people and their relative power (Bramsen et al., 2011; Metcalf & Leake, 2016). Previous studies are limited to standard linguistic features such as the use of phrase "please", frequency of different pronouns or the presence of family-related lexicons.

With the raising of social networks, people have become more connected than ever. At the same time, the language they use has also been changed. When it comes to online chatting, the boundaries between writing and speaking become fuzzy (Crystal, 2001). Abbreviations, repeated characters, emojis and other non-standard spelling variations are more prevalent as they are used to accommodate the invisibility of each other's visual and audio gestures. I hypothesize that these new linguistic features do not only function as a reflection of self-identity or emotion, but they could be a clue to uncovering the complex nature of people's social relationships supported by previous sociolinguistic research which demonstrates that non-standard language can serve as a marker of 'in-group' and 'out-group' identity (Giles & Marlow, 2011; Labov, 1972). However, it has never been fully utilized in the current NLP models.

Conversational AI is of growing importance since it enables easy interaction interface between humans and computers. Researchers are trying to build a system to make it behave more like a human by incorporating human information such as emotion (Zhou et al., 2018), personality (Li et al., 2016; Qian et al., 2018); or even an abstract ability such as empathy (Rashkin et al., 2018). Another aspect that has not been looked into in the literature is the context of social relations. People do not act in isolation but are part of pairs, groups and communities. One could act politely to maintain social harmony and avoid social conflict (Holmes & Wilson, 2013) while acting strongly when he/she is in a higher situational power (Fairclough, 2013).

As spelling variation could be a gold mine to extract information on social relationships, it could be leveraged in a conversational system to make it adaptable to a different degree of closeness as humans are.

In this work, I would like to study the correlation between the spelling variation of the internet language and social relationships and aim to use it to build a dialogue system that can reflect its relationship with the conversational partner. This work will collaborate with Dr Attapol Rutherford, Chulalongkorn University, Thailand.

Propose of the study

This work aims to study linguistic features that can be used to identify people's relationships in conversational texts with ultimate goal to build a conversational model that can use spelling variation to reflect its identity and social relationship with the conversational partner.

Social relationships in focus

To the best of my knowledge, there are no comprehensive categories of social relationships. One sociolinguistic theory explains the idea of social relationship as relative power that operates at the level of individuals in relation to specific others within groups or communities (Fairclough, 2013). It can be explained in two dimensions where horizontal positioning relates to closeness and related constructs such as positive regard, trust and commitment. In contrast, vertical positioning relates to authority and related constructs such as approval and respect among community individuals.

Extended to the relative power concept, I can define the social relationship with two values; degree of closeness and degree of authority. In this way, it could avoid the fuzziness of definition by a vague term such as friends, family, life partner, or colleagues as the relationship is continuum. People usually have different degrees of closeness to their friends but still call their relationship as "friendship".

In this study, there are 5 relationships in focus.

1. Love partner
2. Family
3. Friend
4. Work colleague
5. Online stranger

Throughout the experiment, the social relationships will be referred with a general term (e.g. friend) and two numberical values; closeness and authority. The participant will be asked to indicate the degree of closeness and authority in the relationship and **privately** report to the researchers if it is part of the task.

Please note that, in this study, I only focus on the relative power between people in the conversation caused by their social position. The effect of situational power will be ignored (for example, the relative power due to the relationship between a patient and a doctor, the power affects the speakers only in a health-related situation).

Data collection

Data will be collected from 3 settings representing 3 different perspectives of social relationships; donation, wizard-of-oz and rating settings. The first two represent what people themselves think about their relationship in actual and controlled environments, while the last one represents others' perceptions of the relationship.

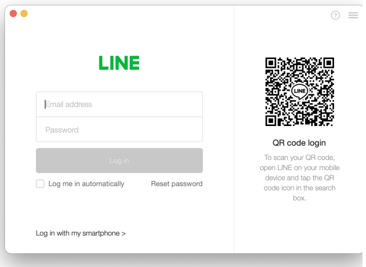
Setting 1: Donation

I will post the instructions on Facebook asking people to donate their exported messages from LINE – a popular messaging application in Thailand. The donor will be asked to specify the degree of closeness and degree of authority to the conversational partner. The data will be anonymized to remove personal information manually by the donor and the researchers. Messages from a specific date or time could be removed by the donor. Sensitive/personal terms (names, addresses, phone numbers) will be replaced by unidentifiable tokens. Researchers will manually verify the resulting data one more time before analyzing it.

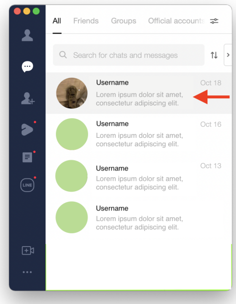
Donation instructions

The post will be a series of images explaining step-by-step instructions to export messages from LINE, anonymize the data and submit it to the researcher. Please see the instruction examples below. The graphic and description will be changed and translated into Thai later.

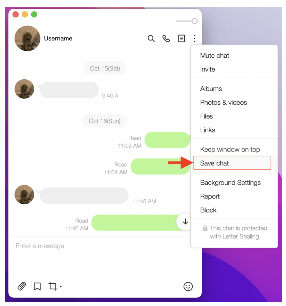
1. Login LINE on your laptop or PC
2. Select a conversation you would like to donate
3. Tap the menu icon at the top of the chat screen and tab **Save chat**.
4. Open the downloaded text file with your preferable word editor.
5. Select a specific date/time to remove and replace any sensitive terms with a pre-defined set of tokens.
6. Lastly, submit the final file to the study's MS Form.



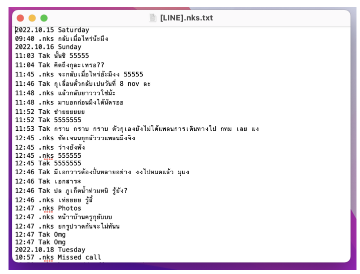
Step 1 Login LINE on your laptop or PC



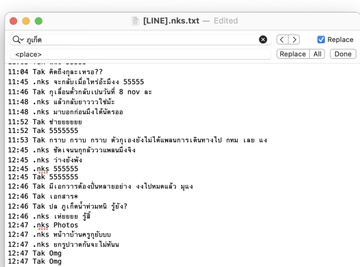
Step 2 Select a conversation you would like to donate



Step 3 Tap the menu icon at the top of the chat screen and tab **Save chat**



Step 4 Open the downloaded text file with your preferable word editor.



Step 5 Select a specific date/time to remove and replace any sensitive terms

The data submission form consists of a detailed explanation of the study's aims, consent form, and task-related questions asking donor's email (for tracking right-to-withdraw), degree of closeness and degree of authority to the conversation partner in the submitting file and, lastly, a file upload button.

Data Anonymization

The donor will be informed to pre-anonymize the data before submitting it. The process could be done either.

* Remove messages from specific dates/times that the donor felt uncomfortable sharing.

* Replace a personal term such as people's name, and address with a predefined set of tokens. I followed the entity types from (Pilán et al., 2022), as followed.

1. <PERSON_[idx]> for Names of people, including nicknames/aliases, usernames and initials
2. <CODE_[idx]> for Numbers and identification codes, such as social security numbers, phone numbers, passport numbers or license plates
3. <LOC_[idx]> for Places and locations, such as cities, areas, countries, addresses, named infrastructures
4. <ORG_[idx]> for Names of organizations, such as public and private companies, schools, universities, public institutions etc.
5. <DEM_[idx]> for Demographic attributes of a person, such as native language, descent, heritage, ethnicity, job titles, ranks, education, physical descriptions, diagnosis, birthmarks, ages
6. <DATETIME_[idx]> for Description of a specific date and time.
7. <QUANTITY_[idx]> for Description of a meaningful quantity, e.g. percentages or monetary values.

Where [idx] is the index of the entity. For example,

```
Bob: WHAT u doin!! ADAM
Adam: I did nothin. Crazy Frank did!!
Bob: He will pay for this
```

The expected fully anonymize will be as followed.

```
Bob: WHAT u doing!! <PERSON_1>
Adam: I did nothin. Crazy <PERSON_2> did!!
Bob: He will pay for this
```

Please note that it is not mandatory to annotate all entities presented in the list, but only entities that are confidential or could be used to trace back to the donor's identity.

Setting 2: Wizard-of-Oz

10-20 recruited participants will be paired and asked to act as if they are in a given role in the relationship while having a conversation about a random seeding topic (also provided by the researcher). The role will be specified with the degree of closeness and authority, and it could be changed if the participant does not have experience related to that role in the relationship. The pairing could be done randomly or chosen by participants if they both agree.

The conversation will be made in LINE application under the researcher's supervision. They will talk in a group chat created between the pair of participants and the researcher. The researcher could intervene if one participant acts inappropriately. At the end of the conversation, the data will be reviewed by the researcher in terms of quality (they are expected to have at least 30 turns), and sensitivity of the content.

Seeding topics

The seeding topics are based on (Godfrey & Holliman, 1993)

1. Weather Climate
2. Care of the Elderly
3. Vacation Spots
4. Gun Control
5. Air Pollution
6. Music
7. Universal Public Service
8. Crime
9. Credit Card
10. Aids
11. Books and Literature
12. Clothing and Dress
13. Camping
14. Movies
15. Right to Privacy
16. Taxes
17. Houses
18. Women's Roles
19. Pets
20. Family Finance
21. Family Life
22. TV Programs
23. Exercise And Fitness
24. Hobbies And Crafts
25. Child Care
26. Recipes/Food/Cooking
27. News Media
28. Job Benefits
29. Capital Punishment
30. Buying a Car
31. Elections and Voting
32. Baseball
33. Fishing
34. Drug Testing
35. Soviet Union

36. Recycling
37. Football
38. Computers
39. Restaurants
40. Politics
41. Choosing a College
42. Basketball
43. Public Education
44. Gardening
45. Painting
46. Vietnam War
47. Federal Budget
48. Middle East
49. Immigration
50. Universal Health Insurance
51. Auto Repairs
52. Metric System
53. Golf
54. Social Change
55. Family Reunions
56. Consumer Goods
57. Home Repairs
58. Space Flight and Exploration
59. Magazines
60. Boating and Sailing
61. Ethics in Government
62. Woodworking

Setting 3: Rating

I will collect public text conversations from Twitter. Then, I will ask 10-20 participants to indicate the degree of closeness and authority perceived from a given dialogue. Each degree will be annotated on 1-3 scale as shown in the the screenshot below.

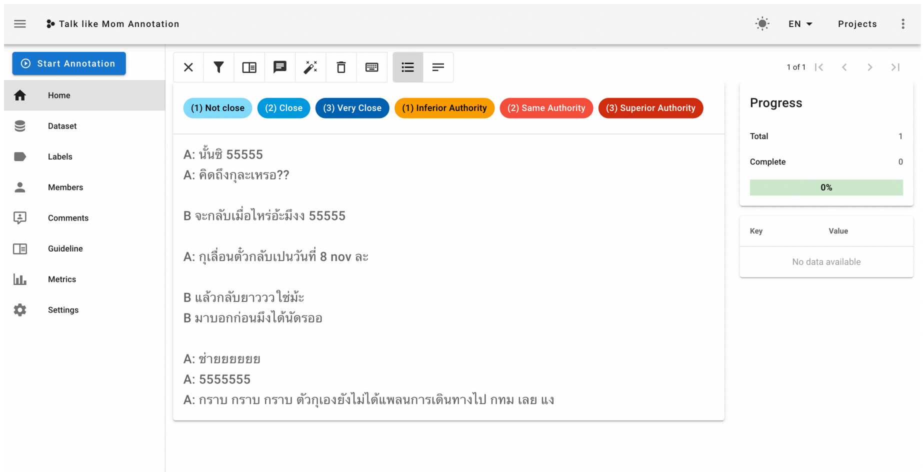


Figure 1: Doccano (Nakayama et al., 2018) will be used as an annotation tool

Data Selection

I will crawl 10,000 Thai tweets via Twitter API. The data will be automatically filtered by number of turns (must have at least 3 turns) and selected by a presence of politness terms such as polite particles, and person-referring expressions (Bilmes, 2001) to maximize the variety of the social relationship. The selected data will have even proportion between dialogue which has the politness terms in standard form, in mispelling form and random selection.

Lastly, the dialogue will be manually selected by the researcher one more time to remove sensitive topics such as sexual acts, graphic violence, symbol or image intented to spread hate based on someone's identity and content related to or encouraged to self-harm.

Linguistic Analysis

In the analyzing stage, the resulting data will be analyzed with the Linguistic Inquiry and Word Count (LIWC) framework (Francis & Booth, 1993), together with the frequency of different patterns of misspellings. The correlation between word counts and degree of closeness and authority will be calculated.

Then, a conversational model will be trained with the same methodology proposed by (Rashkin et al., 2018). BERT will be replaced by WangchanBERTa (Lowphansirikul et al., 2021) as it is trained specifically on Thai corpus. In this work, I will focus on retrieval-based architecture. The based model will be fine-tuned on general dialogue in Thai from Twitter. It will then have a second fine-tuning on our corpus to minimize the negative log-likelihood of selecting the correct candidate.

In the technical detail, the model is given a large set of candidates Y and selects the best response y^* corresponding to the previous utterances x. The response will be scored by $h_x \cdot h_y$ where h_x and h_y are the final embeddings from a language model.

Both automated metrics and human evaluation are employed as an automated metric can give a quick evaluation of the quality of the model, but it does not always correlate with human judgments of dialogue quality (Liu et al., 2016). BLEU score against the actual response will be used as an automated evaluation. For human evaluation, participants will be given a model's output and asked to rate in terms of *Relevance*: did the responses seem appropriate to the conversation? *Fluency*: could you understand the responses? Did the language seem accurate? And *Context*: did the model act corresponding to the given social relationship setting?

Participants

Our main target participants are native Thai-speaking people, as I am interested in studying subtle paralinguistic features which might be difficult for non-natives to understand.

Age, gender identity, and other participants' demographic will not be stored as I only focus on the linguistic features in the conversational text and the social relationship. Recruitment of the experiments will be open regardless of age and gender identity.

References

Bilmes, L. (2001). *Sociolinguistic aspects of thai politeness*. University of California, Berkeley.

Bramsen, P., Escobar-Molano, M., Patel, A., & Alonso, R. (2011). Extracting social power relationships from natural language. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 773–782.

Crystal, D. (2001). *Language and the internet*. Cambridge University Press.

Fairclough, N. (2013). *Language and power*. Routledge.

Francis, M., & Booth, R. J. (1993). Linguistic inquiry and word count. *Southern Methodist University: Dallas, TX, USA*.

Gilbert, E., & Karahalios, K. (2009). Predicting tie strength with social media. *Proceedings of the SIGCHI conference on human factors in computing systems*, 211–220.

Giles, H., & Marlow, M. L. (2011). Theorizing language attitudes existing frameworks, an integrative model, and new directions1. *Annals of the International Communication Association*, *35*(1), 161–197.

Godfrey, J. J., & Holliman, E. (1993). *Switchboard-1 release 2 ldc97s62*. https://catalog.ldc.upenn.edu/LDC97S62

Holmes, J., & Wilson, N. (2013). An introduction to sociolinguistics (4th edi).

Labov, W. (1972). *Language in the inner city: Studies in the black english vernacular*. University of Pennsylvania Press.

Li, J., Galley, M., Brockett, C., Spithourakis, G. P., Gao, J., & Dolan, B. (2016). A persona-based neural conversation model. *arXiv preprint arXiv:1603.06155*.

Liu, C.-W., Lowe, R., Serban, I. V., Noseworthy, M., Charlin, L., & Pineau, J. (2016). How not to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation. *arXiv preprint arXiv:1603.08023*.

Lowphansirikul, L., Polpanumas, C., Jantrakulchai, N., & Nutanong, S. (2021). Wangchanberta: Pretraining transformer-based thai language models. *arXiv preprint arXiv:2101.09635*.

Metcalf, K., & Leake, D. (2016). A computational method for extracting, representing, and predicting social closeness. *Proceedings of the Twenty-second European Conference on Artificial Intelligence*, 1176–1184.

Nakayama, H., Kubo, T., Kamura, J., Taniguchi, Y., & Liang, X. (2018). doccano: Text annotation tool for human [Software available from https://github.com/doccano/doccano]. https://github.com/doccano/doccano

Pilán, I., Lison, P., Øvrelid, L., Papadopoulou, A., Sánchez, D., & Batet, M. (2022). The text anonymization benchmark (tab): A dedicated corpus and evaluation framework for text anonymization. *arXiv preprint arXiv:2202.00443*.

Qian, Q., Huang, M., Zhao, H., Xu, J., & Zhu, X. (2018). Assigning personality/profile to a chatting machine for coherent conversation generation. *Ijcai*, 4279–4285.

Rashkin, H., Smith, E. M., Li, M., & Boureau, Y.-L. (2018). Towards empathetic open-domain conversation models: A new benchmark and dataset. *arXiv preprint arXiv:1811.00207*.

Zhou, H., Huang, M., Zhang, T., Zhu, X., & Liu, B. (2018). Emotional chatting machine: Emotional conversation generation with internal and external memory. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*(1).